Computers, Materials & Continua

Tech Science Press

https://doi.org/10.32604/cmc.2026.084278

REVIEW

A Comprehensive Review of Rating Imputation in Recommender Systems: From Data Completion to Inference-Oriented Missing-Data Estimation

Yong Zheng*

Center for Decision Making and Optimization, School of Computing, College of Tech Futures, Illinois Institute of Technology, Chicago, IL, USA

*Corresponding Author: Yong Zheng. Email: yzheng66@illinoistech.edu

Received: 19 April 2026; Accepted: 15 July 2026; Published: 15 September 2026

ABSTRACT: Recommender systems can alleviate information overload by producing item recommendations tailored to user preferences. The performance usually relies on rich user-item interaction data; however, missing entries introduce sparsity that substantially degrades performance. Early work primarily treated rating imputation as a preprocessing mechanism for mitigating sparsity and alleviating cold-start issues through explicit matrix completion. More recently, missing-data estimation has evolved beyond static preprocessing toward broader inference-oriented paradigms, including pseudo-label estimation, counterfactual inference, and debiasing mechanisms integrated directly into the learning objective. In this paper, we present a structured review of rating imputation and inference-oriented missing-data estimation methods in recommender systems, organized through a conceptual perspective rather than a single formal framework. We organize existing approaches into four stages: statistical approaches, machine learning-based techniques, methods leveraging auxiliary information, and recent advances grounded in causal inference and neural approaches (e.g., deep learning and large language models). Beyond sparsity reduction, we emphasize the role of missing-data estimation in emerging contexts, including correcting selection bias under Missing Not At Random data, inferring partial preferences in multi-criteria recommendation systems, and integrating heterogeneous rating sources. To address an underexplored aspect of the literature, we further analyze potential limitations of imputation and inference-oriented estimation, such as covariance distortion, overconfidence, poisoning effects, and robustness degradation under inaccurate pseudo-labels. By introducing a structured taxonomy, this survey provides a systematic and conceptually organized perspective on missing-data handling in recommender systems and highlights the importance of rigorous bias analysis and robustness evaluation frameworks for developing reliable and unbiased recommendation systems.

KEYWORDS: Recommender system; missing data; rating imputation; causal inference; bias correction

1 Introduction

Recommender systems (RSs) have emerged as effective tools for mitigating information overload, among which collaborative filtering (CF) remains one of the most successful paradigms. However, CF faces several fundamental challenges, notably data sparsity [1,2], the cold-start problem [3,4], and grey-sheep users [5–7] with unusual preferences. For example, the performance of CF is fundamentally limited by data sparsity and the cold-start problem, as users typically provide feedback on only a small subset of available items [1,8,9]. To address these issues, rating imputation has been widely adopted as a preprocessing strategy. By estimating and completing missing interactions, imputation yields a dense pseudo-rating matrix

that facilitates the discovery of latent user-item relationships and, consequently, improves recommendation accuracy [1,10–12].

Although early studies primarily treated rating imputation as a preprocessing technique for mitigating data sparsity through explicit matrix completion, the role of missing-data estimation has expanded substantially in modern recommender systems. In practice, observed ratings rarely satisfy the Missing At Random (MAR) assumption; instead, they more commonly follow a Missing Not At Random (MNAR) mechanism because users are inherently more likely to rate items they prefer [13,14]. This observation has motivated the development of inference-oriented frameworks that estimate missing supervisory signals, pseudo-labels, or counterfactual outcomes to correct selection bias. Representative examples include Error-Imputation-Based (EIB) and Doubly Robust (DR) estimators in causal recommendation frameworks [15,16]. Although these approaches are not strictly equivalent to traditional preprocessing-oriented matrix completion, they remain conceptually related because both paradigms aim to infer missing preference information from partially observed feedback. Beyond bias correction, imputation and missing-preference estimation techniques have also been widely applied in multi-criteria recommender systems (MCRS) to reconstruct incomplete user preferences when only a subset of criteria is observed [17,18]. Moreover, these techniques have been utilized to integrate heterogeneous rating lists from multiple providers by reducing discordance among them [19]. Despite extensive studies on sparsity and debiasing, rating imputation has not been systematically reviewed as a similar paradigm. To fill this critical gap, we present a comprehensive and structured review with a dedicated focus on rating imputation in RSs. We trace its development across four distinct paradigms: early heuristic and statistical methods, machine learning approaches, techniques leveraging auxiliary information, and recent advances based on causal inference and neural approaches. Importantly, we reposition imputation from a simple remedy for data sparsity to a complex, data-driven inference task. These categories discussed in this survey are organized according to their dominant modeling perspectives and historical evolution rather than as strictly independent or mutually exclusive groups. In practice, substantial overlap exists among these paradigms. For example, auxiliary-information-based methods may incorporate machine learning or deep learning models, while modern causal inference frameworks are frequently combined with neural architectures and multimodal representations. Accordingly, the proposed taxonomy is intended to provide a structured conceptual organization of missing-data handling strategies in recommender systems rather than a rigid categorical separation.

In RSs, *prediction*, *imputation*, and *causal inference* are closely related but conceptually distinct paradigms. Rating prediction generally refers to estimating unknown user preferences as a popular recommendation task. Classical rating imputation, in contrast, typically functions as a preprocessing mechanism that explicitly fills missing entries in the user-item matrix to construct a dense pseudo-rating matrix prior to recommendation. In this process, predictive models could be used for the purpose of imputation, but it is still a preprocessing step in RSs. Meanwhile, many modern causal inference and debiasing frameworks do not directly complete the matrix. Instead, they embed missing-preference estimation within the training objective by estimating pseudo-labels, pseudo-errors, or counterfactual supervisory signals associated with unobserved interactions to mitigate selection bias under MNAR settings. Some of these methods additionally incorporate propensity estimation to model the exposure mechanism underlying observed interactions under MNAR settings.

Accordingly, this survey adopts a broader missing-data inference perspective that encompasses both traditional rating imputation methods for explicit matrix completion and modern inference-oriented missing-data estimation frameworks that estimate pseudo-labels, pseudo-errors, or counterfactual supervisory signals from partially observed interactions without necessarily constructing dense rating matrices. We emphasize that this perspective is intended as a conceptual lens for organizing methodologically diverse

approaches under a common missing-data view, rather than a unified formal or algorithmic framework that subsumes them. The connection among these paradigms is conceptual, i.e., they share the goal of inferring missing preference information rather than a claim of mathematical equivalence. In addition, we provide an analysis of often-overlooked risks associated with imputation, including covariance distortion, over-confidence, and poisoning effects [20,21], and highlight the need for rigorous bias analysis and robustness evaluation frameworks. By introducing a structured taxonomy, this survey aims to provide a conceptually organized and systematic perspective on rating imputation, supporting the development of more robust and unbiased recommender systems.

The remainder of this paper is organized as follows. Section 2 introduces preliminary concepts regarding general approaches for missing data imputation, and explores missing data mechanisms and specific recommendation scenarios requiring rating imputation. Section 3 reviews the evolution of imputation methodologies, encompassing heuristic and statistical methods, machine learning and neighborhood-based techniques, auxiliary information-driven imputation, causal inference and debiased imputation, and recent advances in deep learning and large language models. Section 4 critically analyzes the potential risks, bias propagation, and robustness testing associated with imputation. Section 5 outlines open challenges and future research directions. Finally, Section 6 concludes the paper.

2 Missing Data and Imputation in RSs

In this section, we first introduce the issue of missing data in general data science, then discuss the paradigms of missing ratings in RSs, and point out the scenarios requiring rating imputation in RSs.

2.1 Missing Data in General

Before examining rating imputation in RSs, it is important to consider the missing-data problem from a broader data science perspective. Missing values are pervasive across many empirical domains and often prevent the direct application of standard statistical and machine learning methods [22,23]. If not properly addressed, missing data may reduce statistical power, distort covariance structures, and introduce biased parameter estimation and unreliable inference [23,24].

Formally, let $Y \in \mathbb{R}^{n \times p}$ denote a data matrix composed of observed entries Y_{obs} and missing entries Y_{miss}. Following Rubin's classical framework [25], missing-data mechanisms are commonly categorized into Missing Completely At Random (MCAR), Missing At Random (MAR), and Missing Not At Random (MNAR) [26]. Under MCAR, missingness is independent of both observed and unobserved data. Under MAR, the probability of missingness depends only on observed variables. In contrast, MNAR arises when missingness is directly related to the missing values themselves [24]. Correctly understanding these mechanisms is essential because the effectiveness and validity of imputation strategies strongly depend on the underlying missingness assumptions.

Methods for handling missing data can generally be grouped into four categories: deletion methods [22,27], single imputation methods [28,29], likelihood-based estimation [24,30], and multiple imputation methods [31,32].

- **Deletion Methods:** Deletion-based approaches remove incomplete observations either globally or locally [22,27]. Although computationally simple, these methods often result in severe information loss and biased estimation when the MCAR assumption does not hold [23].
- **Single Imputation Methods:** Single imputation replaces each missing entry with a single estimated value [28,29]. Common examples include mean substitution, regression imputation, and classification-based estimation. These methods restore data completeness but typically ignore uncertainty in the

imputed values. Therefore, they may distort variance and correlations, leading to overly confident statistical conclusions [22,23].

- **Likelihood-Based Estimation:** Likelihood-based approaches, e.g., expectation-maximization and full information maximum likelihood estimation, estimate model parameters directly from incomplete observations without explicitly filling missing entries [24,30]. By treating missing values as latent variables, these methods properly account for uncertainty and can produce unbiased parameter estimates under the MAR assumption [23].
- **Multiple Imputation Methods:** Multiple imputation methods generate multiple completed datasets to explicitly model uncertainty in missing values [31,32]. Each completed dataset is analyzed independently, and the results are subsequently combined using standard pooling procedures [24]. Compared with single imputation, these approaches generally provide more reliable statistical inference under MAR assumptions [23]. However, they introduce higher computational cost and still rely on missingness assumptions that may not hold in real-world behavioral data [26].

Missing-data handling and imputation have also attracted increasing attention in related machine learning domains, including incomplete graph learning, where missing attributes and incomplete graph structures substantially affect downstream graph representation and prediction tasks [33]. Although these missing-data mechanisms and handling strategies originate from classical statistics, they are highly relevant to recommender systems because user-item interaction data is inherently incomplete and often exhibits extreme sparsity and non-random missingness. In particular, MNAR plays a central role in recommendation environments because users selectively interact with items according to personal preferences and platform exposure dynamics. The recommender-system-specific interpretation of MCAR, MAR, and MNAR is further discussed in Section 2.2.

2.2 Rating Imputation in RSs

Although the general imputation techniques discussed in Section 2.1 provide a solid theoretical foundation, their direct application to RSs is constrained by the distinctive characteristics of user–item interaction data. In conventional data analysis, missing rates of 20% to 30% are already considered severe and require careful treatment. In contrast, RSs operate on highly sparse rating matrices, where missing rates frequently exceed 95% and may reach 99% in large-scale industrial settings [8,34]. For example, in movie platforms such as Netflix or e-commerce platforms such as Amazon.com, there are millions of items on these platforms, but each user may rate a limited set of items in their history, which results in much higher sparsity in the rating matrix. Under such extreme sparsity, classical statistical imputation methods often become computationally infeasible or yield highly biased estimates.

2.2.1 Paradigms of Missing Ratings

Following Rubin's classical framework, missingness in RSs can be categorized into MCAR, MAR, and MNAR paradigms.

To better discuss missing ratings in RSs under different missing-data paradigms, Table 1 summarizes the characteristics of MCAR, MAR, and MNAR using representative examples from both educational settings and recommender systems. Although these mechanisms originate from classical missing-data theory, their practical implications differ substantially in recommendation environments due to user self-selection behavior, interaction sparsity, and platform exposure dynamics. Under the MCAR assumption, missing ratings are independent of both observed and unobserved preferences, which may occur when interactions are randomly lost because of logging failures or temporary system outages. In MAR scenarios, missingness depends on observed variables rather than the missing preference itself; for example, users

with low activity levels or limited browsing histories may be less likely to provide ratings. In contrast, MNAR is generally regarded as the dominant missing-data mechanism in recommender systems because users selectively rate items they strongly prefer or dislike while ignoring neutral or uninteresting items. Consequently, missing interactions often contain informative latent preference signals rather than purely random omissions, which fundamentally distinguishes recommendation-oriented imputation from conventional statistical missing-data completion and motivates the development of debiasing-oriented, causal, and inference-based recommendation frameworks. More specifically, under MNAR settings, observed ratings no longer represent an unbiased sample of user preferences because the exposure and selection processes are inherently preference-dependent. As a result, recommendation models trained directly on observed interactions may overestimate popular or highly exposed items while underrepresenting unobserved preferences. Accordingly, many modern debiasing and causal recommendation frameworks can be interpreted as imputation-oriented mechanisms that attempt to estimate missing supervision signals, pseudo-errors, or counterfactual outcomes for unobserved interactions.

Table 1: Comparison of missing-data mechanisms in education and recommender systems.

Mechanism	Characteristics	Example in Education (Missing Grades)	Example in RSs
MCAR	Missingness is independent of both observed and unobserved variables	Student grades are missing due to accidental database corruption or random system failure	Ratings are randomly lost because of logging failures or temporary platform outages
MAR	Missingness depends only on observed variables	Students with lower attendance rates are less likely to complete optional assignments or surveys	Users with low activity levels or limited browsing history are less likely to provide ratings
MNAR	Missingness depends directly on the missing values themselves	Students with poor academic performance intentionally avoid reporting grades	Users selectively rate items they strongly like or dislike while ignoring neutral or uninteresting items

When ratings are MCAR or MAR, the probability that an entry is missing is independent of its unobserved true value. For example, under MCAR, ratings may be missing due to random system logging failures or interface issues that affect all items uniformly. Under MAR, missingness may depend on observed factors such as user activity level or item exposure; for instance, less active users may provide fewer ratings overall, while the likelihood of a rating being missing remains unrelated to the user's true preference for a specific item. In such cases, the missingness mechanism can be ignored, and analyses based solely on observed data produce unbiased parameter estimates and predictions. Accordingly, early collaborative filtering methods and heuristic imputation techniques implicitly adopted the MAR assumption, treating missing entries primarily as a sparsity issue to be completed. For example, neighborhood-based CF methods [35,36], such as KNN-based CF, compute user-user or item-item similarity directly from observed ratings, often ignoring missing entries or treating them as implicit zero contributions during similarity calculation. In such formulations, the absence of a rating is not interpreted as a systematic signal but is instead handled as missing information to be smoothed over when constructing neighborhoods and generating

predictions. Similarly, matrix factorization techniques learn latent representations of users and items based on observed ratings [37].

Empirical evidence indicates that the MAR assumption rarely holds in real-world RSs. Interaction data arises from voluntary user behavior rather than controlled experiments, where users tend to rate items they prefer or frequently encounter while ignoring those they dislike or have not engaged with [13,14]. This self-selection process results in an imbalance in which high ratings dominate observed data, whereas low ratings are underrepresented and more likely to be missing. Such patterns imply that the probability of missingness depends on the underlying rating value, which is consistent with the MNAR mechanism. The consequences of MNAR are substantial. Observed ratings no longer represent the true preference distribution across items. Applying MAR-based methods, such as standard matrix factorization or naive mean imputation, without accounting for this bias introduces significant selection effects. Thus, learned preferences become distorted, which often leads to over-recommendation of popular items and degraded recommendation performance, particularly in terms of top-k accuracy across the full item space. Therefore, modern imputation strategies must explicitly address non-random missingness. The focus has shifted from matrix densification toward the use of imputation as a principled tool for bias correction and unbiased learning.

These distinct missingness mechanisms have also shaped the evolution of recommendation paradigms. Early heuristic and neighborhood-based approaches primarily treated missing ratings as a sparsity problem under implicit MCAR or MAR assumptions, where imputation mainly served as a matrix densification strategy. By contrast, modern causal inference and debiasing frameworks explicitly model MNAR behavior arising from user self-selection and exposure bias. In these settings, imputation is no longer limited to explicit rating completion, but instead functions as an inference-oriented mechanism for estimating pseudo-labels, pseudo-errors, or counterfactual supervisory signals associated with unobserved interactions. Recent deep learning and LLM-based approaches further extend this paradigm by leveraging multimodal semantic representations to infer missing preference information from unstructured data sources.

2.2.2 Imputation Scenarios

Building on this foundation, we summarize the application scenarios that require specialized imputation strategies in RSs, as described in the following.

Data Sparsity and the Cold-Start Problem. In this setting, missing interactions are primarily treated as incomplete preference observations caused by extreme sparsity rather than systematic exposure bias. Accordingly, imputation mainly serves as a matrix densification mechanism to reconstruct latent user-item relationships.

Extreme sparsity in the user-item interaction matrix represents a primary scenario that necessitates imputation. In real-world applications, users interact with only a small fraction of the available catalog, which leads to missing rates that often exceed 95% to 99%. Under such conditions, traditional CF algorithms, such as user-based KNN [38] and item-based KNN [39] approaches, are unable to identify sufficient overlap in interactions, where user-based methods suffer from a lack of co-rated items between users and item-based methods lack sufficient common users who have rated both items, which renders similarity estimation unreliable. In addition to neighborhood-based methods, model-based CF approaches, such as latent factor models, are also affected by extreme sparsity. Early implementations, including conventional SVD [40,41], require a fully observed rating matrix, which makes imputation a necessary preprocessing step to enable training. By contrast, later matrix factorization methods [37,42] are designed to operate directly on sparse data by optimizing only over observed entries. Although this formulation avoids the need for explicit

imputation, the severe lack of observations still limits the ability to learn reliable latent representations, often leading to suboptimal generalization in highly sparse settings.

The problem becomes more serious during the "cold-start" phase [3,43], where new users or items lack adequate historical interactions. Imputation mitigates this limitation by constructing a dense pseudo-rating matrix. By estimating unobserved interactions using demographic information or auxiliary attributes, it introduces sufficient overlap between users and items, thereby enabling neighborhood formation and supporting personalized recommendations in sparse settings.

The primary advantage of data imputation in this scenario is its ability to mitigate the severe sparsity inherent in recommender systems. By estimating missing entries and converting a highly sparse user-item matrix into a dense pseudo-rating matrix, imputation restores the overlap required for effective interaction modeling. Despite these benefits, naive imputation introduces notable risks. From a computational perspective, converting a matrix with extreme sparsity (e.g., 99% missing) into a dense representation substantially increases memory usage and computational cost [34]. From a statistical perspective, poorly estimated values, such as those obtained through zero or mean substitution, can introduce noise, distort the covariance structure, and reduce variance [12,34]. In addition, recent work identifies the risk of "poisonous imputation", where inaccurate estimates deviate from the true values and undermine the debiasing objective, which leads to degraded model performance [21].

Correcting Selection Bias and Non-Random Missingness. Unlike sparsity-oriented recommendation settings, missing interactions under MNAR are strongly associated with user exposure, self-selection, and platform recommendation dynamics. In this context, imputation shifts from simple rating completion toward counterfactual estimation and debiasing-oriented inference. As implied by the MNAR mechanism, disregarding the non-random nature of missing data introduces substantial selection bias, which distorts learned user preferences and degrades recommendation performance, particularly in terms of top-k accuracy [14,15]. To mitigate this issue, the recommendation community has developed several debiasing strategies. Generative approaches attempt to jointly model the missingness mechanism and the rating process through causal graphs; however, constructing accurate graphs in complex real-world settings remains challenging [21,44]. Alternatively, Inverse Propensity Scoring (IPS) reweights observed samples to correct distributional shifts, yet it is highly sensitive to propensity estimation and often exhibits high variance under extreme sparsity [21,44].

Within this context, data imputation provides a direct mechanism for bias correction. By estimating prediction errors for unobserved ratings or assigning plausible pseudo-labels, such as lower ratings for missing interactions, imputation adjusts the observed data distribution toward the underlying global distribution [44]. This principle underlies Error-Imputation-Based (EIB) estimators. In addition, imputation constitutes a key component of Doubly Robust (DR) estimators, which combine IPS and EIB in the same framework [15]. A notable advantage of imputation-based debiasing, particularly within DR, is the double robustness property, where unbiased estimation is achieved if either the imputation model or the propensity model is correctly specified [21]. Furthermore, incorporating imputation reduces the variance associated with IPS alone, which stabilizes optimization and improves the bias-variance trade-off [44].

Despite these advantages, imputation-based debiasing introduces several risks. The effectiveness of EIB and DR estimators depends critically on the quality of the imputation model [45]. When pseudo-labels are mis-specified or poorly calibrated, especially under simplistic modeling assumptions, the resulting errors propagate through training and limit the effectiveness of bias correction [20,44]. More critically, recent work identifies the phenomenon of "poisonous imputation", where imputed values deviate substantially from the ground truth, which increases both bias and variance and ultimately degrades model performance [21].

Partial Preference Completion in MCRS. In MCRS settings, missingness often occurs at the sub-criteria level rather than the overall interaction level. Accordingly, imputation focuses on recovering incomplete multidimensional preference structures while preserving inter-criteria dependencies. Traditional recommender systems rely on a single overall rating to represent user preferences. In contrast, multi-criteria recommender systems extend this paradigm by allowing users to evaluate items across multiple distinct criteria [46–49]. This formulation is widely adopted in practical applications, including platforms such as TripAdvisor for hotel recommendations [50,51], Yahoo! Movies for film evaluation [52], and OpenTable for dining experiences [53].

One of the challenges in MCRS is the "partial preference" problem, which results in structurally incomplete criteria-level ratings [54]. This missingness arises from two primary sources. First, during explicit data collection, requiring users to rate all predefined criteria introduces cognitive burden and user fatigue. To mitigate this issue, platforms often request ratings for only a subset of criteria, leaving the remaining entries unobserved. Second, in modern systems, multi-criteria ratings are frequently inferred from unstructured textual reviews through natural language processing techniques [55–58]. Since user-generated reviews rarely cover all criteria comprehensively, the extracted profiles remain incomplete [59].

To address partial preference, two main methodological directions have been explored. One line of work develops specialized models, such as tensor factorization and probabilistic graphical approaches, which operate directly on sparse observations without explicit completion [60]. However, these methods often incur high computational cost and face difficulties in capturing inter-criteria dependencies under extreme sparsity. The alternative, and more commonly adopted, strategy introduces a dedicated pre-recommendation imputation stage to reconstruct complete utility profiles for each user-item pair [18].

Imputation provides a practical advantage by densifying the data, which enables standard multi-criteria models to better capture dependencies among criteria and improves both coverage and accuracy [17]. A complete profile allows the system to model user preferences in a more coherent manner. Nevertheless, imputation introduces potential risks. Simple heuristic methods, such as mean substitution or standard Naive Bayes, assume independence across criteria and fail to reflect realistic perceptual relationships, for example the interaction between cleanliness and perceived service quality [18]. Such assumptions can introduce noise and distort the underlying preference structure. To address these limitations, recent approaches employ sequential models, including Recurrent Naive Bayes, to capture inter-criteria dependencies in a structured manner [18], or leverage contextual semantics from LLM-based representations such as BERT to infer missing criteria from textual reviews [17]. Although these methods improve estimation accuracy, they introduce additional computational overhead and require careful calibration to prevent error propagation during recommendation generation [17,18].

Integration of Heterogeneous Rating Resources. In heterogeneous recommendation environments, different rating resources may follow fundamentally different missingness mechanisms. For example, explicit ratings are often influenced by user self-selection, whereas implicit feedback is strongly affected by exposure and interaction frequency. As a result, unified imputation across heterogeneous resources requires careful consideration of modality-specific missingness assumptions. Ratings are increasingly used as standardized indicators across domains such as ESG evaluation, academic journal assessment, and healthcare benchmarking. In these scenarios, decision-makers often aggregate ratings from multiple independent providers to form the final evaluation. However, each provider typically evaluates only a subset of subjects, which introduces substantial structural missingness in the combined matrix.

A similar scenario in RSs refers to the area of meta-recommendation systems that combine multiple sources to produce a unified output [61–63]. For example, a movie recommendation system may integrate ratings from platforms such as IMDb, Rotten Tomatoes, and Metacritic. Similarly, healthcare decisions may

rely on combined hospital evaluations from providers such as Medicare, LeapFrog, and U.S. News. Since each provider typically evaluates only a subset of items due to differences in scope or resource constraints, the resulting provider-item matrix exhibits substantial structural missingness [19].

It is important to distinguish this scenario from traditional recommendation settings. While conventional ratings represent individual user-item interactions, the "ratings" in meta recommendations represent institutional or platform-level evaluations, effectively forming a Provider-Item matrix rather than a User-Item matrix. Thus, integrating these heterogeneous lists aligns with the concept of a meta-recommendation system, which blends multiple independent recommendation sources into a unified output [61]. This fundamentally differs from cross-domain recommender systems, which typically focus on transferring learned knowledge or user interaction patterns from a data-rich source domain to a different, sparse target domain [64,65].

Traditional approaches attempt to infer the underlying scoring mechanisms or identify the key explanatory factors within each rating system [19]. However, this strategy becomes impractical when combining multiple sources, as the underlying criteria and scoring schemes differ across providers. As a result, this approach requires extensive auxiliary information and customized modeling for each source, which limits scalability and general applicability [19]. In this setting, data imputation shares the same mathematical objective as collaborative filtering and matrix completion [19]. It provides an efficient alternative by estimating missing entries using only the observed ratings within the combined matrix, thereby enabling an evaluation framework across providers. However, imputing heterogeneous rating lists introduces distinct challenges. Providers often exhibit conflicting assessments for the same items, referred to as "upsets", and employ different rating scales and distributions. Simple imputation strategies, such as averaging, fail to reconcile these discrepancies and may distort the aggregated outcomes [19]. Consequently, this scenario requires specialized methods, including discordance minimization based on quadratic programming, which incorporate rank-based relationships and penalize inconsistencies across providers to achieve reliable imputation [19].

2.2.3 Related Surveys

Rating imputation is closely related to several major research directions in RSs, including sparsity reduction, debiasing recommendation, causal recommendation, and LLM-based recommendation systems. As illustrated in Fig. 1, imputation-oriented mechanisms lie at the intersection of these research areas because many recommendation paradigms implicitly or explicitly estimate missing preference information for unobserved interactions. Prior surveys in these related domains frequently discuss imputation-related techniques as part of broader recommendation challenges.

In sparsity-oriented recommender-system surveys, imputation is commonly introduced as one possible strategy for alleviating incomplete user-item interactions and cold-start problems. However, sparsity reduction extends substantially beyond imputation itself and also includes matrix factorization, latent factor learning, graph propagation, transfer learning, and representation learning approaches [37,66]. Similarly, debiasing recommendation surveys often discuss pseudo-label estimation and counterfactual correction mechanisms related to imputation, particularly under MNAR settings [15,21]. Nevertheless, many debiasing methods rely on alternative strategies such as inverse propensity weighting, exposure modeling, reweighting, and unbiased risk estimation rather than explicit missing-value estimation [67,68].

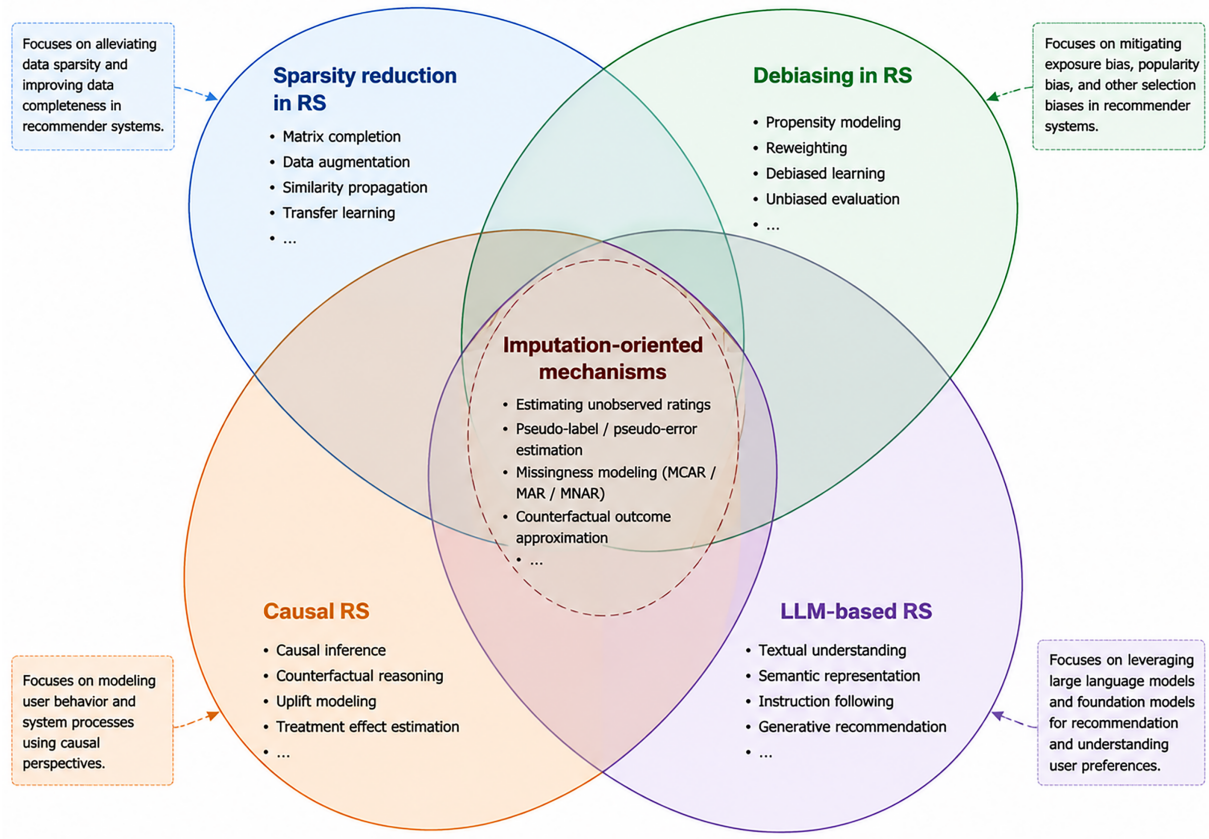

Figure 1: Relationship between imputation-oriented recommendation research and related directions.

A similar relationship exists in causal recommendation research. Although several causal inference frameworks perform implicit imputation through counterfactual estimation or pseudo-error approximation, the broader causal recommendation literature also includes treatment effect estimation, intervention modeling, uplift modeling, and causal representation learning that do not directly rely on imputation-oriented mechanisms [69,70]. Likewise, recent LLM-based recommender-system surveys discuss semantic preference inference, review understanding, and generative recommendation [71–73]. However, many LLM-driven recommendation paradigms focus on reasoning, conversational interaction, or generation-oriented recommendation tasks that extend beyond missing-data estimation.

Therefore, although imputation-oriented recommendation mechanisms overlap substantially with these broader research directions, existing surveys typically discuss imputation only as one component within larger recommendation frameworks. To the best of our knowledge, there remains no dedicated survey that systematically organizes rating imputation itself as a missing-data handling paradigm spanning statistical completion, auxiliary-information integration, causal debiasing, and deep semantic inference in recommender systems.

3 Imputation Methodologies

Having identified the scenarios in which data imputation is essential, the discussion turns to the methodologies used to generate these substitute values. As noted in previous sections, the effectiveness of the subsequent recommendation phase depends critically on the quality of imputed data, since inaccurate estimates can distort the covariance structure and amplify existing biases. This dependency has driven the continuous evolution of imputation strategies within the recommender systems community.

This section provides a comprehensive review of the evolution of imputation in recommender systems across five methodological paradigms. The literature was collected from major academic databases, including Google Scholar, IEEE Xplore, ACM Digital Library, and ScienceDirect. The search process employed combinations of keywords related to rating imputation and missing-data handling in recommender systems, including "rating imputation", "missing data in recommender systems", "matrix completion", "data sparsity imputation", "missing not at random", "counterfactual imputation", and "imputation-based debiasing". The review primarily focused on peer-reviewed conference and journal publications published between 2001 and early 2026, and we finally included 31 unique publications in our analysis. Papers were included if they addressed missing interaction handling through explicit or implicit imputation-oriented inference mechanisms, such as rating completion, pseudo-label estimation, pseudo-error estimation, or counterfactual missing-data modeling. Foundational missing-data studies and representative recommender-system surveys were also included to establish the statistical and methodological context of the field. Pure recommendation optimization or exposure modeling methods without an imputation-related component were excluded from the primary taxonomy. Rather than exhaustively enumerating all existing studies, this survey emphasizes representative and influential works to provide a structured overview of the evolution of imputation techniques in recommender systems.

The paradigms discussed in the following subsections include heuristic and statistical filling techniques in Section 3.1, predictive machine learning and neighborhood-based approaches in Section 3.2, methods leveraging auxiliary information in Section 3.3, causal inference frameworks for debiasing under MNAR settings in Section 3.4, and recent neural approaches, including deep learning and LLM-based architectures, in Section 3.5. These developments illustrate the transition of recommender systems from traditional preprocessing-oriented matrix completion toward more inference-oriented, bias-aware, and semantically informed missing-data estimation frameworks.

These methodological categories are organized primarily according to their dominant modeling perspectives and historical evolution rather than as strictly independent or mutually exclusive paradigms. In practice, substantial overlap exists among these categories. For example, auxiliary-information-based approaches may incorporate machine learning or deep neural architectures, while modern causal inference frameworks are frequently combined with representation learning and multimodal models. Accordingly, the taxonomy presented in this section is intended to provide a structured conceptual organization of missing-data handling strategies rather than a rigid categorical separation.

To further illustrate the representative methodological progression discussed in this survey, Fig. 2 conceptually summarizes the evolution of imputation paradigms based on the representative methods reviewed in Tables 2–6 in the following subsections. The figure highlights the transition from early statistical and neighborhood-based matrix completion approaches toward auxiliary-information-driven methods, causal debiasing frameworks, and recent deep learning and LLM-based paradigms.

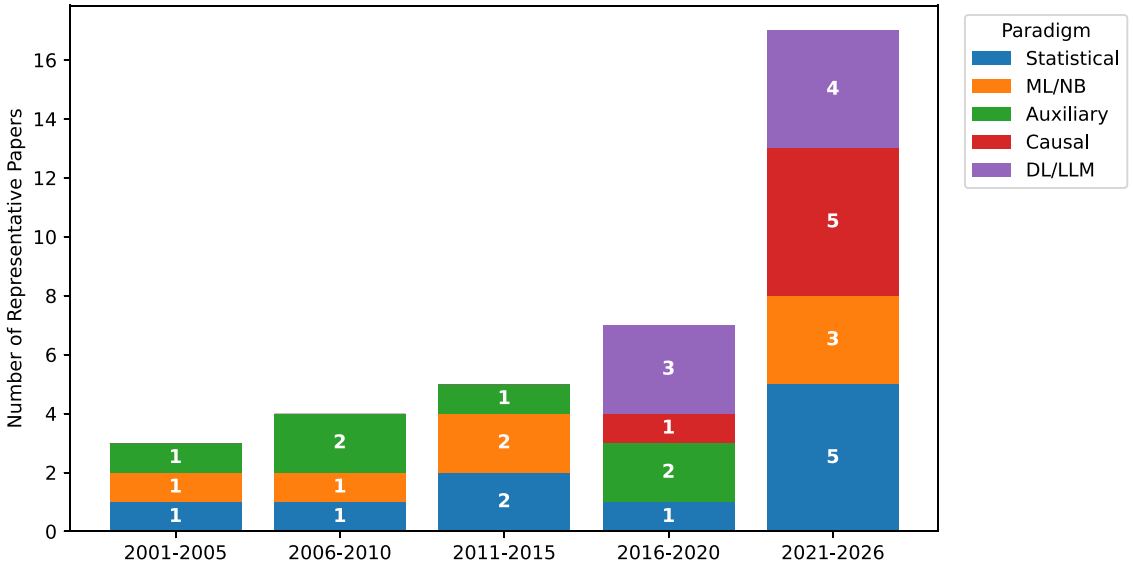

Figure 2: Conceptual evolution of representative imputation paradigms in recommender systems.

Table 2: Summary of heuristic and statistical imputation methods.

Methods	Imputation Type	Scenarios & References
Zero/Null Imputation	Single Imputation	Sparsity [74,75]
Mean/Average Imputation	Single Imputation	Sparsity [2,76], Bias Reduction (MNAR) [10], Partial Preference (MCRS) [18], Heterogeneous Rating Lists [19]
Median Imputation	Single Imputation	Sparsity [75,77]
Mode Imputation	Single Imputation	Sparsity [74]
Min/Max Imputation	Single Imputation	Sparsity [2]
Bayesian Multiple Imputation	Multiple Imputation	Sparsity, Bias Reduction (MNAR) [10]
Chained Equations	Multiple Imputation	Sparsity, Partial Preference (MCRS) [17]
Expectation-Maximization	Likelihood-based Estimation	Sparsity [78]

Table 3: Summary of machine learning and neighborhood-based imputation methods.

Methods	Imputation Type	Scenarios & References
Naive Bayes	Single Imputation	Sparsity [11], Partial Preference (MCRS) [18]
Support Vector Machines	Single Imputation	Sparsity [11]

(Continued)

Table 3 (continued)

Methods	Imputation Type	Scenarios & References
Decision Trees & Random Forest	Single Imputation	Sparsity [11], Partial Preference (MCRS) [17], Heterogeneous Rating Resources [19]
K-Nearest Neighbors	Single Imputation	Sparsity [79]
Auto-Adaptive Imputation	Single Imputation	Sparsity [80]
Recurrent Naive Bayes	Single Imputation	Partial Preference (MCRS) [18]
EBMI	Multiple Imputation	Bias Reduction (MNAR) [10]
EM-SVD	Likelihood-based Estimation	Sparsity [78]

Table 4: Summary of auxiliary information-driven imputation methods.

Methods	Imputation Type	Scenarios & References
Item Attributes	Single Imputation	Sparsity, Cold-Start [81]
User Demographics (e.g., gender)	Single Imputation	Sparsity, Cold-Start [82]
Social Trust Networks	Single Imputation	Sparsity, Cold-Start [83–85]
Auxiliary-Enhanced Multiple Imputation	Multiple Imputation	Bias Reduction (MNAR) [86]

Table 5: Summary of causal inference and debiased imputation methods.

Methods	Imputation Type	Scenarios & References
Error Imputation-Based Methods	Single Imputation	Bias Reduction (MNAR) [15]
Doubly Robust (DR) Estimator	Single Imputation	Bias Reduction (MNAR) [15,87]
Joint Learning DR (DR-JL)	Single Imputation	Bias Reduction (MNAR) [15]
AutoDebias	Single Imputation	Bias Reduction (MNAR) [87]
Conservative/Calibrated DR (CDR, DCE)	Single Imputation	Bias Reduction (MNAR & Instability) [20,21]
Inductive Bias-Relaxed DR (UDR, UIDR)	Single Imputation	Bias Reduction (User/Item-specific Bias) [45]
CounterCLR	Single Imputation	Bias Reduction (Confounding & MNAR) [88]

Table 6: Summary of deep learning and LLM-based imputation methods.

Methods	Imputation Type	Scenarios & References
MLPs	Single Imputation	Sparsity, Cold-Start [89]
Graph Neural Networks	Single Imputation	Missing Modalities, Sparsity [90]
BERT-based Semantic Imputation	Single Imputation	Sparsity, Partial Preference (MCRS) [17]
LLM Prompting & Fine-tuning	Single Imputation	Sparsity [91]
Denoising Autoencoders	Multiple Imputation	Sparsity [92]
Deep Generative Models	Multiple Imputation	Bias Reduction (MNAR) [93]
Deep Latent Variable Models	Likelihood-based Estimation	Sparsity [94]

3.1 Heuristic and Statistical Imputation

Early imputation approaches in recommender systems primarily treated missing user-item interactions as an incomplete data problem caused by extreme sparsity. These methods relied on fundamental statistical estimation and deterministic filling strategies to reconstruct missing entries in the rating matrix. Although these methods are simple, they establish fundamental statistical assumptions that continue to influence modern imputation strategies. As summarized in Table 2, these methods can be broadly categorized into single imputation, multiple imputation, and likelihood-based estimation, each corresponding to different recommendation scenarios, including extreme sparsity and bias mitigation.

Foundational single imputation methods are widely adopted to mitigate data sparsity by replacing missing entries with heuristic estimates. For example, zero or null imputation assigns a minimum value, such as zero, to all unobserved interactions, which densifies the matrix with minimal computational cost [74,75]. Mean imputation is among the most commonly used strategies. In addition to addressing general sparsity [2,76], it serves as a baseline for partial preference completion in MCRS [18] and is applied to resolve inconsistencies in heterogeneous rating resources [19]. Early studies have also explored mean substitution for bias reduction in MNAR settings [10].

However, mean imputation exhibits notable limitations. It tends to distort the empirical distribution by creating an artificial concentration around the mean, which reduces variance and weakens correlations among variables. In addition, the imputed values are often non-integer and may not align with the discrete rating scales used in practice. To address these issues, alternative strategies based on median [75,77] and mode [74] have been proposed. Median imputation provides greater robustness to outliers and better reflects the central tendency under skewed distributions, while mode imputation preserves the discrete nature of rating scales by assigning the most frequently observed value.

Beyond these approaches, min and max imputations have been adopted in specific scenarios with distinct motivations [2]. Assigning minimum values can encode implicit negative feedback under MNAR assumptions, where unobserved interactions are interpreted as lack of interest. Conversely, maximum values may be used to capture strong positive signals in neighborhood-based estimation or to select the most optimistic predictions among candidate estimates. In addition, min and max values are often employed as boundary constraints to ensure that predicted ratings remain within valid ranges.

To address the statistical noise and variance distortion associated with single-point estimates, the community has explored multiple imputation and likelihood-based estimation. Bayesian Multiple Imputation generates several plausible values for each missing entry and has been applied to reduce selection bias while supporting sparse neighborhood-based filtering [10]. Chained Equations have demonstrated effectiveness in MCRS by iteratively estimating missing sub-criteria through conditional modeling [17]. Likelihood-based

approaches, including the expectation-maximization algorithm, have also been integrated with SVD-based frameworks to estimate missing values by maximizing the expected likelihood of observed data [78].

Despite their statistical foundation, heuristic and statistical imputation methods exhibit both advantages and limitations. Their primary strength lies in computational efficiency and ease of implementation, as they enable the construction of dense matrices that support memory-based collaborative filtering and standard matrix factorization without architectural changes. However, these methods typically rely on the assumption that data follows MCAR or MAR mechanisms. As a result, applying uniform imputation across large portions of missing data lacks personalization. More importantly, naive statistical filling, particularly mean substitution, can distort covariance structures, reduce variance, and introduce substantial noise under MNAR conditions. These limitations motivate the transition from heuristic approaches to predictive machine learning models that can capture non-linear relationships and user-specific preferences.

3.2 Machine Learning and Neighborhood-Based Imputation

To overcome the limitations of statistical filling strategies, subsequent research increasingly reframed missing-data imputation as a personalized preference prediction problem. Rather than relying solely on global statistical assumptions, these approaches exploit user-item interaction patterns, similarity structures, and predictive learning models to infer missing ratings. As summarized in Table 3, this category includes both neighborhood-based methods and machine learning-driven approaches that estimate missing preferences from observed behavioral patterns under sparse recommendation settings.

A representative framework in this paradigm is imputation-boosted collaborative filtering, which employs predictive classifiers such as Naive Bayes, support vector machines, and decision trees to construct a high-quality pseudo-rating matrix before applying Pearson correlation-based collaborative filtering [11]. In this formulation, rating imputation is cast as a classification problem, where discrete rating levels (e.g., 1 to 5 stars) are treated as class labels. The input features are typically derived from the users' observed ratings on other items, while in highly sparse settings, demographic attributes such as age, gender, occupation, and postal code can be incorporated as auxiliary features to support personalized prediction [11]. Beyond addressing general sparsity, tree-based models such as random forest, have been adopted as robust baselines for resolving opinion inconsistencies in heterogeneous rating resources [19] and for completing partial preferences in MCRS [17].

In parallel, neighborhood-based imputation methods exploit local similarity structures. K-nearest neighbors estimates missing values by aggregating preferences from similar users or items [79]. However, applying global neighbors to impute all missing entries across the entire matrix is computationally intensive and may accumulate prediction errors. This limitation motivated the development of Auto-Adaptive Imputation (AutAI) [80]. Instead of performing global imputation, AutAI identifies a task-specific neighborhood tailored to the active user and the target item. It restricts imputation to the intersection of related users, defined by shared rating histories, and related items that are co-rated with the target item. Grounded in nearest-neighbor information principles, this strategy focuses on imputing only the most informative missing entries. As a result, AutAI reduces computational cost while limiting the propagation of noisy pseudo-ratings into subsequent prediction stages [80].

Machine learning approaches have also been extended to capture structural dependencies in multi-criteria data through refined feature representations. While standard Naive Bayes assumes conditional independence, sequence-aware models such as recurrent Naive Bayes model inter-criteria dependencies by formulating imputation as a sequential prediction task. In this setting, the feature space evolves dynamically, where the overall item rating provides initial context and previously imputed sub-criteria are progressively incorporated as additional features for subsequent predictions [18].

The methods described above operate as single imputation techniques, where they produce a single deterministic estimate for each missing entry. Although this design supports computational efficiency, it introduces a key statistical limitation. The use of a single point estimate ignores prediction uncertainty, which results in overconfident imputations and an underestimation of variance [11].

To address this limitation, researchers have explored the integration of multiple imputation and likelihood-based estimation within machine learning frameworks. Instead of producing a single estimate, multiple imputation generates multiple plausible rating matrices that are analyzed jointly. For example, Extended Bayesian Multiple Imputation (EBMI) has been incorporated into imputed neighborhood-based CF, where repeated sampling from a posterior distribution preserves data variability and mitigates bias induced by MNAR data [10]. Likelihood-based methods, such as Maximum Likelihood estimation via the Expectation-Maximization algorithm, can also be applied in principle along with the SVD approach [78]. However, these approaches are less practical in large-scale recommendation settings, as iterative likelihood maximization over highly sparse matrices incurs substantial computational cost compared to direct multiple imputation or efficient single-imputation classifiers.

Despite notable improvements in personalization and accuracy over simple statistical methods, predictive machine learning and neighborhood-based approaches exhibit inherent limitations. Their effectiveness depends on the availability of sufficiently dense local interaction data to train models or identify reliable neighbors. In extreme cold-start scenarios, where new users or items lack historical interactions, these methods cannot extract meaningful features or construct valid neighborhoods. This limitation necessitates the incorporation of auxiliary information beyond the rating matrix to support robust recommendation.

3.3 Auxiliary Information-Driven Imputation

Although machine learning and neighborhood-based approaches improve personalized preference estimation, their dependence on observed user-item interactions makes them highly vulnerable to extreme sparsity and cold-start conditions. When a new user or item enters the system without any historical interactions, neither feature extraction from observed ratings nor reliable neighbor identification is feasible. To address this limitation, the imputation paradigm has been extended to incorporate external auxiliary information. As summarized in Table 4, classical auxiliary-driven imputation approaches primarily rely on single imputation and auxiliary-enhanced multiple imputation to construct a dense semantic representation for cold-start entities.

A direct strategy for alleviating extreme sparsity is to incorporate explicit content and demographic metadata into single imputation frameworks. Item genres and product attributes enable the estimation of missing ratings by exploiting a user's affinity for related categories [81]. In a similar manner, demographic-based imputation utilizes user profiles, including age, gender, occupation, and geographic location, to infer preferences for unrated items by aggregating signals from comparable demographic groups [82]. Beyond these metadata, relational auxiliary data, particularly social trust networks, have been incorporated into imputation models. Under the principle of homophily, explicit trust relationships, especially bidirectional connections, facilitate the identification of reliable neighbors for estimating missing ratings [83,84]. To further refine these initial estimates, matrix factorization methods such as probabilistic matrix factorization [84] and non-negative matrix factorization [85] have been adapted to constrain the latent factors of a target user to align with those of trusted connections, thereby mitigating sparsity and cold-start effects.

Furthermore, auxiliary information has been employed to address non-random missingness. From a statistical perspective, missing ratings are often correlated with their unobserved values, which leads to the MNAR problem. To mitigate this issue, auxiliary variables that are not part of the primary model but are strongly associated with either the ratings or the missingness process, such as user demographics or temporal

interaction patterns, are incorporated into the imputation framework [23,86]. By conditioning on these variables, the dependence between missingness and unobserved ratings is reduced, which shifts the effective data-generating process closer to a MAR assumption and alleviates selection bias [23]. Within this setting, auxiliary-enhanced multiple imputation leverages these variables to define a posterior predictive distribution for missing entries. Multiple stochastic samples are drawn to construct m completed rating matrices, each of which is analyzed independently and subsequently combined to produce the final prediction [86]. This integration enables partial bias correction while preserving the uncertainty associated with missing data [86].

Despite these advantages, auxiliary information-driven imputation presents two fundamental limitations that motivate further methodological developments. On one hand, although auxiliary data expands the information space and alleviates sparsity, it does not explicitly model the causal mechanisms underlying user exposure and item selection. Achieving unbiased recommendation therefore requires counterfactual reasoning, which has led to the adoption of causal inference models as discussed in Section 3.4. On the other hand, these methods are primarily limited to structured metadata and simple relational information, and lack the capacity to capture complex semantic preferences from unstructured data sources such as textual reviews, images, or heterogeneous knowledge graphs. Addressing this limitation has driven the transition toward deep learning and LLM-based approaches as described in Section 3.5.

3.4 Causal Inference and Debiased Imputation

To address the limitations of conventional sparsity-oriented imputation and explicitly account for non-random missingness, RS research has increasingly shifted toward causal inference and debiasing frameworks. Unlike traditional imputation approaches that primarily treat missing entries as incomplete preference observations, causal models interpret missing interactions as outcomes of user exposure, selection behavior, and recommendation feedback loops associated with the MNAR mechanism [67,68]. These approaches extend imputation beyond explicit matrix completion toward inference-oriented estimation of counterfactual supervisory signals, such as pseudo-labels and pseudo-errors, for unobserved interactions. As summarized, modern causal imputation frameworks primarily focus on debiased learning and counterfactual estimation to mitigate selection bias in recommender systems.

As shown in Table 5, these methods rely on generating pseudo-labels or pseudo-errors for unobserved interactions to approximate the unobserved counterfactual outcomes. In essence, this mechanism is conceptually aligned with traditional rating imputation, as both aim to infer missing signals for unobserved entries; however, causal methods reformulate this process as error or label imputation within the learning objective rather than explicitly completing the rating matrix. Here, the term "label" refers to the supervision signal used for training, such as the relevance indicator or the observed rating value associated with a user-item interaction, which is extended to unobserved pairs through imputation to enable unbiased estimation.

The foundational approach in this paradigm is the error imputation-based method, which imputes missing labels and trains the prediction model using both observed ratings and imputed pseudo-labels [15,45]. To relax the unbiasedness requirement of these error imputation-based methods, the Doubly Robust (DR) estimator was introduced. By combining an error imputation model with an inverse propensity scoring model, DR achieves theoretical unbiasedness when either the propensity estimates or the imputed pseudo-errors are accurate [15,45]. Here, the propensity model estimates the exposure probability of observed interactions under the MNAR mechanism rather than directly reconstructing missing ratings. Motivated by this property, a range of advanced imputation strategies has been developed. For example, DR-JL refines the imputation model by jointly minimizing prediction errors on observed data [15,21]. To reduce reliance on heuristic tuning, AutoDebias incorporates a small set of unbiased uniform data through meta-learning to guide the generation of imputed errors [87]. More recent work addresses instability caused by inaccurate

pseudo-labels. Methods such as Conservative DR (CDR) and Doubly Calibrated Estimator (DCE) filter high-variance imputed values and calibrate error probabilities to improve robustness [20,21]. To address the limitations of imputation, inductive bias-relaxed DR methods, including User-DR (UDR) and User-Item-DR (UIDR), have been proposed. These approaches learn a constrained propensity model that ensures unbiasedness even when the imputed pseudo-labels deviate from the true labels due to user-specific and item-specific inductive biases [45]. Moreover, counterfactual imputation frameworks such as CounterCLR employ contrastive learning on non-random missing data to align representations of exposed and unexposed interactions [88].

From a statistical perspective, these causal models operate as single imputation techniques. In the DR formulation, the imputation model produces a single deterministic pseudo-error, such as $\hat{e}_{u,i}$, for each unobserved user-item pair to support gradient-based optimization [15,21]. Although multiple imputation could theoretically capture uncertainty in MNAR settings by approximating a MAR distribution while preserving variance [23], its direct application is computationally prohibitive. Constructing and training m dense rating matrices at large scale requires excessive memory and computation. Instead, recent methods approximate the benefits of multiple imputation through stochastic neural techniques. For instance, CDR employs Monte Carlo dropout to estimate both the mean and variance of pseudo-errors, which provides an efficient approximation of posterior sampling without explicitly generating multiple datasets [21].

Causal imputation methods provide a rigorous foundation for debiased recommendation, yet they exhibit both strengths and limitations. Their primary advantage lies in the theoretical guarantee of unbiasedness and the ability to separate exposure effects from intrinsic user preferences [45]. However, their effectiveness depends critically on the quality of imputed pseudo-labels. Applying DR-based imputation indiscriminately across all unobserved pairs can introduce unreliable estimates that degrade performance [21,44]. In addition, computing pseudo-errors and propensity scores over the entire unobserved space incurs substantial computational cost. Moreover, extracting reliable counterfactual representations or accurate propensity estimates from complex and unstructured data sources, including text, images, and knowledge graphs, remains challenging for traditional causal formulations. These limitations have motivated the adoption of deep learning and LLM-based approaches for robust representation learning in recommendation.

3.5 Deep Learning and Large Language Models

Although auxiliary-information-driven and causal inference frameworks substantially improve sparsity handling and debiasing, many of these approaches still rely on structured metadata, handcrafted features, or simplified statistical assumptions. These approaches face inherent limitations in capturing complex, nonlinear semantic patterns from unstructured modalities, including large-scale textual reviews, raw images, and heterogeneous knowledge graphs. To address this limitation, recent research has increasingly incorporated deep learning and LLM-based models to perform more expressive and semantically informed imputation, as summarized in Table 6.

The evolution of deep imputation methods began with adapting neural architectures to reconstruct missing interactions. Lee et al. employed Variational Autoencoders (VAEs) to learn latent user and item representations, followed by Multi-Layer Perceptrons (MLPs) to infer missing ratings from these representations [89]. With the rise of multimodal recommendation, the objective extended beyond rating completion to the reconstruction of missing modality features. By leveraging the structural properties of user-item interactions, graph neural networks have been applied to infer missing visual and textual features through graph-based interpolation on item-item relations [90].

In recent years, generative AI, particularly LLMs, has gained widespread adoption across domains such as education [95–97], finance [98–100], and healthcare [101,102], enabling advanced data generation and reasoning capabilities. These developments have also influenced recommender systems [72,103], where generative models are increasingly used to extract structured signals from unstructured data and to enhance representation learning under sparse and incomplete settings.

In parallel, pre-trained language models have been utilized to extract structured signals from unstructured text. For example, in MCRS, BERT-based semantic imputation derives fine-grained sub-criteria ratings directly from user reviews, thereby enriching the preference representation [17]. More recently, imputation has been formulated as a generative task, thereby reframing missing value estimation as conditional generation rather than explicit matrix completion. For example, LLM-based models, such as GPT-2, are fine-tuned with structured prompts to predict missing values by leveraging pre-trained knowledge [91]. From a statistical perspective, these approaches operate as single imputation methods, since they produce a single deterministic estimation for each missing entry.

To capture uncertainty in unobserved data and address the MNAR mechanism, deep learning has also been integrated with multiple imputation and likelihood-based estimations. For instance, denoising autoencoders have been extended to generate multiple imputed datasets, such as MIDA [92], which are subsequently combined to preserve data variability. In the context of selection bias, deep generative models explicitly model the joint distribution of data and missingness, enabling debiased multiple imputation under MNAR conditions [93]. Alternatively, deep latent variable models, such as DeepLTRS, follow a likelihood-based paradigm. By optimizing the evidence lower bound through variational inference, these models estimate the joint distribution of observed ratings and auxiliary signals without explicitly constructing completed datasets [94].

Although most existing deep learning and LLM-based recommendation models [72,103] are not explicitly framed as rating imputation approaches, several emerging paradigms share important conceptual similarities with imputation-oriented inference. For example, sequential [104,105] and generative recommender systems [106,107] frequently estimate future interactions and latent preference trajectories from incomplete behavioral histories, which can be interpreted as implicit preference completion under sparse observations. Similarly, diffusion-based recommendation models iteratively reconstruct latent preference distributions from corrupted or incomplete interaction signals, resembling probabilistic imputation processes. In graph-based recommendation, debiasing and counterfactual graph learning frameworks may also estimate missing exposure relationships and latent supervision signals under MNAR conditions [108].

Recent LLM-based recommender systems further extend recommendation from numerical matrix completion toward semantic preference inference and reasoning-oriented interaction modeling, and have boosted the development of several emerging applications, such as prompt-based rating prediction [72,109], review-to-rating inference [110], semantic profile completion [111], and conversational dialogue simulation [112,113] and preference elicitation [114,115]. Instead of relying solely on observed rating matrices, these approaches leverage semantic reasoning and contextual world knowledge to infer latent user preferences and unobserved interactions. However, these paradigms also introduce new challenges, including hallucinated preferences, semantic inconsistency, unstable pseudo-label generation, and difficulties in evaluating semantically inferred preferences without explicit ground-truth ratings. Future research may increasingly reinterpret recommendation-oriented missing-data handling as a broader inference and representation learning problem rather than purely numerical matrix completion.

3.6 Summary

As summarized in Table 7, the five imputation paradigms differ substantially in their underlying assumptions, strengths, limitations, and suitable application scenarios, providing an explicit basis for comparing when each paradigm is most appropriate and what trade-offs it entails. Heuristic and statistical methods offer efficient and interpretable matrix completion, but they rely on simplified missingness assumptions and provide limited personalization. Machine learning and neighborhood-based approaches improve preference estimation by exploiting interaction patterns, yet they remain vulnerable to sparsity and cold-start conditions. Auxiliary-information-driven methods mitigate this limitation by incorporating side information, although their effectiveness depends heavily on the availability and quality of external features. Causal inference and debiased imputation methods are better suited to MNAR settings because they explicitly address exposure and selection bias, but they are sensitive to propensity estimation errors and unstable pseudo-labels. Deep learning and LLM-based approaches further extend imputation through multimodal representation learning and semantic inference, but they introduce substantial computational cost and reduced interpretability. A more explicit comparison across these paradigms reveals several recurring trade-offs. In terms of computational cost, statistical and neighborhood-based methods remain the most efficient, whereas causal DR estimators and LLM-based approaches incur the highest overhead because they estimate pseudo-errors over the entire unobserved space or rely on large-scale generative inference. When it comes to the concerns of robustness, statistical and early machine learning methods implicitly assume MCAR/MAR and degrade markedly under MNAR, while causal and debiased methods are explicitly designed for MNAR but become sensitive to propensity estimation errors and unstable pseudo-labels. The best choice of the methodologies is largely scenario-driven rather than strictly hierarchical: statistical methods remain competitive under mild sparsity where interpretability and low cost matter, machine learning and auxiliary-information methods are preferable when sufficient interaction history or side information is available, and causal and neural approaches become necessary only when MNAR bias correction or multimodal semantic inference is the dominant concern. Moreover, a direct quantitative performance comparison across paradigms is difficult because reported results are obtained on heterogeneous datasets, sparsity levels, and missingness conditions, and few studies share the same unbiased MAR test protocols; we therefore compare paradigms along qualitative dimensions of cost, robustness, and suitability rather than absolute accuracy. Overall, no single paradigm dominates across all recommendation settings; instead, the appropriate imputation strategy depends on the missingness mechanism, data sparsity level, availability of auxiliary information, and robustness requirements.

Table 7: Comparative analysis of representative imputation paradigms.

Paradigm	Primary Assumption	Strengths	Limitations	Effective Scenarios
Statistical Imputation	Missing entries can be estimated from global statistical patterns under simplified missingness assumptions	Simple implementation, low computational cost, interpretable statistical foundation	Limited personalization capability, weak semantic modeling, sensitive to severe sparsity and MNAR bias	Mild sparsity, preliminary matrix densification, small-scale recommendation settings
ML & NB-based Imputation	User preferences can be inferred from interaction similarity and predictive behavioral patterns	Improved personalization and local preference estimation	Highly dependent on historical interactions, vulnerable to extreme sparsity and cold-start problems	Moderate sparsity with sufficient user-item interaction history

(Continued)

Table 7 (continued)

Paradigm	Primary Assumption	Strengths	Limitations	Effective Scenarios
Auxiliary Information-Driven Imputation	External side information can compensate for missing interactions	Effective cold-start mitigation, incorporation of user and item semantics	Strong dependence on auxiliary feature quality and structured metadata availability	Cold-start recommendation, sparse environments with rich side information
Causal Inference & Debiased Imputation	Missingness is driven by exposure and user selection behavior under MNAR conditions	Explicit debiasing capability, counterfactual estimation, improved robustness against selection bias	Sensitive to propensity estimation errors, training instability, variance amplification	MNAR recommendation settings, exposure bias correction, debiased recommendation learning
Deep Learning & LLM-based Imputation	Complex semantic relationships can be inferred from multimodal and high-dimensional representations	Strong representation learning capability, multimodal semantic inference, flexible generative modeling	High computational and memory cost, reduced interpretability, risk of overfitting under extreme sparsity	Large-scale recommendation systems, multimodal recommendation, semantic preference inference

4 Risks, Bias Analysis, and Robustness Testing

This section shifts the focus from imputation effectiveness to its potential failure modes, which remain underexplored in existing literature. While the previous sections highlight the effectiveness of data imputation in multiple scenarios, it is important to recognize that imputation is not a universal solution. In practice, handling missing data presents inherent trade-offs [44]. Naively filling unobserved interactions, particularly under the highly skewed and complex distributions common in recommender systems, can introduce substantial unintended effects. Moreover, related adversarial recommendation studies, including poisoning attacks and synthetic user manipulation [116,117], further demonstrate that imputation-oriented recommendation frameworks may be vulnerable to maliciously generated interaction signals and biased pseudo-preference estimation. As imputation techniques evolve from simple neighborhood-based heuristics to advanced counterfactual and neural inference models, careful consideration of their statistical limitations, robustness degradation, and algorithmic risks becomes essential. This section provides a critical examination of the limitations of data imputation and outlines key considerations for ensuring model reliability.

4.1 The Dark Side of Imputation

In real-world deployment, attempts to densify highly sparse user-item matrices through deterministic imputation often introduce significant operational risks. A primary concern is the large-scale injection of noise. In highly skewed recommendation datasets, imputation across the entire unobserved space can generate values that substantially deviate from true user preferences. These inaccurate and high-variance pseudo-labels are referred to as "poisonous imputations" [21]. Instead of facilitating unbiased learning, such values distort the optimization objective and introduce substantial noise into the gradient descent process, which ultimately degrades recommendation accuracy.

In addition to algorithmic noise, single imputation introduces critical statistical issues during model training and evaluation. Missing data theory indicates that replacing missing entries with a single predicted value, without incorporating stochastic variability, artificially suppresses the natural variance of the dataset [23]. As a result, the standard errors of model parameter estimates are systematically underestimated,

since the uncertainty across multiple plausible imputations is ignored [23]. This effect leads to statistical overconfidence, where the model treats imputed pseudo-labels as true observations. The predictive uncertainty associated with the MNAR mechanism is neglected. This overconfidence compromises the validity of offline evaluation and produces models that are fragile when exposed to dynamic real-world user behavior.

Importantly, the types and extent of imputation risks differ across recommendation scenarios. Based on the fundamental challenges discussed in previous sections, we highlight the most critical risks that require careful consideration under four representative contexts:

- **Sparsity and Cold-Start:** In this setting, the dominant risk is *poisonous imputation*. When new users or items lack sufficient interaction history, imputation models must extrapolate under severe information scarcity. As a result, deterministic estimates are prone to large deviations from true preferences, which introduces substantial noise into the learning objective and may hinder rather than facilitate cold-start adaptation [21].
- **Partial Preference Completion in MCRS:** For MCRS, the primary concern is *variance underestimation*. Sub-criteria such as food, service, and atmosphere exhibit inherent correlations while maintaining distinct variability. Deterministic imputation suppresses this natural variation and inflates inter-criteria dependencies [23]. Consequently, the model becomes overconfident in imputed relationships and loses the ability to represent nuanced user preferences.
- **Bias Reduction in MNAR:** Under MNAR, the most critical issue is *bias amplification through feedback loops*. If imputation relies on inaccurate propensity estimates or incorrect assumptions about the missingness mechanism, the resulting pseudo-labels inherit existing biases in the observed data. When these biased estimates are iteratively incorporated into model training, selection bias and the Matthew effect are progressively reinforced [44].
- **Integration of Heterogeneous Rating Resources:** When integrating diverse rating signals, such as combining explicit ratings (e.g., 1–5 stars) with implicit feedback (e.g., clicks and views), a key challenge lies in the *mismatch in missingness mechanisms*. These data sources are generated through fundamentally different observational processes; for instance, missing explicit ratings often follow MNAR behavior driven by user self-selection and extreme preferences, whereas missing implicit feedback is largely influenced by exposure bias [19,44]. Applying a unified imputation strategy across such heterogeneous sources without accounting for these differences can introduce spurious correlations and propagate systematic errors throughout the recommendation model.

4.2 Bias Analysis

To systematically diagnose the issues discussed in the previous section and to prevent the amplification of imputation risks through the feedback loop of recommender systems, the research community has developed rigorous bias analysis frameworks that emphasize principled evaluation protocols and sensitivity-based diagnostics. Instead of relying solely on theoretical assumptions, these approaches focus on quantitatively assessing how imputation mechanisms behave under varying degrees of missingness.

An example of this effort is the adoption of unbiased evaluation protocols that address the limitations of conventional testing procedures. Evaluating debiased models on observational holdout data tends to reward the memorization of existing exposure and selection biases, which undermines the validity of offline metrics. To overcome this issue, benchmark datasets such as *Coat* and *Yahoo! R3* have been widely used [15,87]. These datasets contain subsets of collected ratings obtained through randomized item assignment, which removes user-driven selection effects [87]. Under this protocol, models are trained on skewed MNAR data and evaluated on uniformly sampled MAR test sets [15]. This separation enables a more reliable assessment of debiasing capability, independent of bias memorization.

In addition, robust bias analysis requires evaluating the sensitivity of imputation methods to violations of their underlying assumptions. Classical missing data theory shows that treating MNAR data as MAR introduces systematic bias into estimated relationships, with the direction and magnitude determined by the dependence between selection variables and unobserved outcomes [23]. To examine this effect in practice, sensitivity analysis is conducted by modifying the assumed missingness mechanism during evaluation. This process may involve introducing controlled distortions or varying the accuracy of propensity estimates [15]. By tracking changes in metrics such as mean squared error under these variations, researchers can characterize the bias trajectory and identify the conditions under which imputed signals transition from beneficial approximations to harmful poisonous imputations.

4.3 Robustness Testing

To mitigate the algorithmic risks associated with data imputation and to ensure that debiasing mechanisms do not degrade system performance, identifying and measuring bias alone is insufficient. Developing reliable recommender systems requires proactive safeguards and rigorous stress-testing procedures. Accordingly, the research community has established three complementary quality assurance strategies that operate across training, testing, and evaluation stages.

During training, uncertainty-aware filtering and calibration are employed to control the impact of imputed pseudo-labels. Since indiscriminate use of all imputations introduces unreliable signals, modern approaches estimate uncertainty through techniques such as Monte Carlo dropout, which performs multiple stochastic forward passes to compute the mean and variance of each imputed error [21]. By applying adaptive thresholds, the system filters out high-variance imputations and retains only reliable pseudo-labels for gradient-based optimization. In addition, probability calibration methods, including Platt scaling and related techniques, are used to adjust imputed errors and propensity estimates, thereby reducing overconfidence and preventing variance amplification [20].

At the testing stage, semi-synthetic noise injection is used to evaluate robustness under controlled perturbations [45]. For instance, as demonstrated in the experimental protocol of inductive bias-relaxed DR estimators [45], a ground-truth rating matrix is systematically modified to introduce structured distortions into the pseudo-labels. Typical strategies adopted in such protocols include flipping a proportion of specific ratings to extreme values (e.g., the ONE strategy), rotating the rating distribution by shifting high ratings to lower values (e.g., the ROTATE strategy), and adding Gaussian noise proportional to the original ratings (e.g., the SKEW strategy) [45]. Model stability is then assessed using metrics such as relative errors, where these approaches were demonstrated to maintain stable performance and preserve unbiased estimation under these adverse conditions.

Finally, uncertainty-aware multiple imputation addresses the statistical overconfidence associated with single deterministic imputation. Instead of relying on a single completed dataset, multiple imputation generates m distinct datasets by sampling from a posterior predictive distribution, often using Markov Chain Monte Carlo or related techniques. The recommendation model is applied independently to each dataset, and the resulting predictions are combined using Rubin's rules [23]. This procedure restores the natural variance that is suppressed by deterministic imputation and enables the estimation of reliable standard errors and confidence intervals, thereby allowing the system to appropriately reflect predictive uncertainty in practical deployment.

5 Open Challenges and Future Directions

The evolution of data imputation in recommender systems, from basic heuristic filling to advanced causal inference and neural approaches, has significantly enhanced the ability to address sparsity and

MNAR biases. Nevertheless, several open challenges remain, including a substantial gap between theoretical advances and robust deployment in large-scale industrial environments. Building on the limitations and risks identified in previous sections, we outline the following key challenges and corresponding research directions that deserve further investigation.

- **Dynamic Missingness Mechanisms and Temporal Imputation.** Most existing causal and deep imputation models assume a static missingness mechanism. In practice, user preferences, item popularity, and exposure strategies evolve continuously, which renders both selection bias and exposure bias time-dependent [44,87]. Future work should move beyond static matrix completion toward dynamic modeling of interaction trajectories, such as integrating sequential recommendation models with online meta-learning to update propensity estimates and imputation strategies, or incorporating fast online revision algorithms (e.g., sequential SVD updating with exponential decay) to efficiently track non-stationary user tastes and adapt to temporal data drifts.
- **Scalability and Efficiency of Advanced Imputation Models.** Although DR estimators and LLM-based approaches achieve strong predictive performance, they incur substantial computational and memory overhead. For example, DR methods require estimating pseudo-errors for all unobserved pairs, while LLM-based approaches rely on large-scale generative inference [91]. Improving scalability remains essential for practical deployment. Future research should focus on efficient approximations of the unobserved space, including sampling-based strategies for causal models and parameter-efficient techniques such as LoRA, quantization, and knowledge distillation for LLM-based imputation.
- **The Dilemma of Unbiased Evaluation and Learning without MAR Data.** Reliable evaluation of imputation methods typically depends on unbiased MAR datasets, such as Coat and Yahoo! R3. However, collecting such data through random exposure strategies negatively impacts user experience and platform performance, which limits feasibility in real-world systems [44]. An important research direction is to enable unbiased learning and evaluation without relying on explicit MAR data. Emerging approaches include invariant learning and generative modeling techniques that attempt to recover true preferences directly from MNAR observations. Extending these approaches to general recommendation settings remains an open challenge.
- **Multimodal-Driven Preference Inference and Rating Imputation.** The availability of rich side information has created opportunities to infer missing preferences from complex multimodal data, including textual reviews, images, and videos. Prior work has shown that pre-trained models, such as BERT, can infer missing criteria ratings from textual content [17,94]. A related line of work also addresses missing multimodal features themselves, e.g., via graph-based propagation on the item-item graph [90], which is complementary to inferring the preferences those features inform. Future imputation models should integrate semantic signals from these modalities to generate high-quality pseudo-ratings. This objective is challenging, as the translation of unstructured content into numerical ratings may introduce spurious correlations arising from the semantic gap between the two representations. Addressing this challenge motivates the development of multimodal LLM-based approaches that align heterogeneous features with explicit preference representations, together with validation mechanisms that ensure the reliability of the generated pseudo-ratings prior to their incorporation into model training.
- **Imputation in Specialized Recommender Systems.** Despite progress in standard recommendation settings, the application of imputation techniques to specialized domains, such as context-aware recommender systems [118–120] and point-of-interest recommendation [121,122], remains limited. By introducing contextual variables such as time, location, and environmental conditions, these systems extend the traditional user-item matrix into higher-dimensional representations, which intensifies sparsity. Future work should explore imputation solutions to reconstruct missing interactions in

these settings. Developing causal estimators that account for complex, context-dependent exposure mechanisms represents a promising direction for advancing recommendation quality in these domains.

6 Conclusions

This survey has systematically examined the evolution, methodological foundations, and inherent risks of rating imputation, which serves as a central strategy for densifying sparse data and approximating unobserved counterfactual outcomes to support unbiased learning. More broadly, this survey highlights that imputation should be understood not solely as a preprocessing technique, but as a fundamental inference problem under missing and biased data. We position this work as a conceptual reframing of missing value imputation rather than as a universal solution. Given the heterogeneity of missingness mechanisms and application contexts, no single framework can comprehensively address all scenarios, a conclusion further supported by our paradigm-level comparison.

The proposed taxonomy traces the progression of imputation techniques from early heuristic and statistical methods to advanced theoretical frameworks. In particular, causal inference paradigms have redefined the role of imputation by embedding error and label estimation directly within the learning objective, rather than relying on explicit matrix completion [15,21]. In addition, recent advances in neural approaches, such as deep learning and LLM technologies, demonstrate strong capability in extracting latent semantic signals from unstructured multimodal data, such as textual reviews, thereby enabling more accurate inference of missing criteria and the generation of high-quality pseudo-ratings [17,91]. This survey also highlights that imputation introduces nontrivial risks alongside its benefits. Forced completion of missing data may produce high-variance poisonous imputations that disrupt optimization, while deterministic filling strategies suppress natural variance and lead to overconfident and fragile models [23]. Addressing these issues requires moving beyond naive evaluation practices toward rigorous and systematic protocols.

Future research is expected to shift from static and uniform matrix completion toward dynamic, scalable, and context-aware imputation frameworks. Promising directions include relaxing assumptions on imputation accuracy, leveraging generative models for cross-modal validation, and extending imputation methods to address exponential sparsity in multi-criteria and context-aware settings. These developments will support the construction of recommender systems that achieve not only high predictive accuracy but also robustness and reliable uncertainty estimation.

Acknowledgement: Not applicable.

Funding Statement: The author received no specific funding for this study.

Availability of Data and Materials: Not applicable.

Ethics Approval: Not applicable.

Conflicts of Interest: Given his role as Editorial Board Member of this journal, Yong Zheng had no involvement in the peer review of this article and had no access to information regarding its peer review. Full responsibility for the editorial process for this article was delegated to another journal editor. The authors declare no other conflicts of interest.

References

1. Idrissi N, Zellou A. A systematic literature review of sparsity issues in recommender systems. Soc Netw Anal Min. 2020;10(1):15. doi:10.1007/s13278-020-0626-2.
2. Lestari S, Afdila ME, Pratama YA. Imputation missing value to overcome sparsity problems in the recommendation system. J RESTI. 2023;7(6):1285–91. doi:10.29207/resti.v7i6.5300.

3. Lika B, Kolomvatsos K, Hadjiefthymiades S. Facing the cold start problem in recommender systems. Expert Syst Appl. 2014;41(4):2065–73. doi:10.1016/j.eswa.2013.09.005.
4. Volkovs M, Yu G, Poutanen T. Dropoutnet: addressing cold start in recommender systems. Adv Neural Inf Process Syst. 2017;30:4964–73.
5. Zheng Y, Agnani M, Singh M. Identification of grey sheep users by histogram intersection in recommender systems. In: Advanced data mining and applications. Cham, Switzerland: Springer International Publishing; 2017. p. 148–61. doi:10.1007/978-3-319-69179-4_11.
6. Zheng Y, Agnani M, Singh M. Identifying grey sheep users by the distribution of user similarities in collaborative filtering. In: Proceedings of the 6th Annual Conference on Research in Information Technology (RIIT '17); 2017 Oct 4–7; Rochester, NY, USA. p. 1–6.
7. Zheng Y. Using outlier detection to identify grey-sheep users in recommender systems: a comparative study. Comput Mater Contin. 2025;83(3):4315–28. doi:10.32604/cmc.2025.063498.
8. Koren Y, Rendle S, Bell R. Advances in collaborative filtering. In: Recommender systems handbook. New York, NY, USA: Springer; 2022. p. 91–142. doi:10.1007/978-1-0716-2197-4_3.
9. Kluver D, Ekstrand MD, Konstan JA. Rating-based collaborative filtering: algorithms and evaluation. In: Social information access: systems and technologies. Cham, Switzerland: Springer International Publishing; 2018. p. 344–90. doi:10.1007/978-3-319-90092-6_10.
10. Su X, Khoshgoftaar TM, Greiner R. Imputed neighborhood based collaborative filtering. In: Proceedings of the 2008 IEEE/WIC/ACM International Conference on Web Intelligence and Intelligent Agent Technology; 2008 Dec 9–12; Sydney, NSW, Australia. p. 633–9.
11. Su X, Khoshgoftaar TM, Zhu X, Greiner R. Imputation-boosted collaborative filtering using machine learning. classifiers. In: Proceedings of the 2008 ACM Symposium on Applied Computing; 2008 Mar 16–20; Fortaleza, Brazil. p. 949–50.
12. Su X, Khoshgoftaar TM, Greiner R. A mixture imputation-boosted collaborative filter. In: Proceedings of the Twenty-First International Florida Artificial Intelligence Research Society Conference (FLAIRS 2008); 2008 May 15–17; Coconut Grove, FL, USA. p. 312–6.
13. Marlin BM, Zemel RS, Roweis S, Slaney M. Collaborative filtering and the missing at random assumption. In: Proceedings of the Twenty-Third Conference on Uncertainty in Artificial Intelligence; 2007 Jul 19–22; Vancouver, BC, Canada. p. 267–75.
14. Steck H. Training and testing of recommender systems on data missing not at random. In: Proceedings of the 16th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining; 2010 Jul 25–28; Washington DC, USA. p. 713–22. doi:10.1145/1835804.1835895.
15. Wang X, Zhang R, Sun Y, Qi J. Doubly robust joint learning for recommendation on data missing not at random. In: Proceedings of the 36th International Conference on Machine Learning; 2019 Jun 9–15; Long Beach, CA, USA. p. 6638–47.
16. Wang X, Zhang R, Sun Y, Qi J. Combating selection biases in recommender systems with a few unbiased ratings. In: Proceedings of the 14th ACM International Conference on Web Search and Data Mining; 2021 Mar 8–12; Virtual. p. 427–35. doi:10.1145/3437963.3441799.
17. Göksel G, Aydın A, Batmaz Z, Kaleli C. A novel missing value imputation for multi-criteria recommender systems. Inf Sci. 2025;712:122139. doi:10.1016/j.ins.2025.122139.
18. Rismala R, Novia Wisesty U, Sthevanie F. Recurrent Naive Bayes for multi-criteria recommender systems: a novel approach for partial preference imputation. Interdiscip J Inf Knowl Manag. 2026;21:3. doi:10.28945/5696.
19. Park YW, Kim J, Zhu D. Discordance minimization-based imputation algorithms for missing values in rating data. Mach Learn. 2024;113(1):241–79. doi:10.1007/s10994-023-06452-4.
20. Kweon W, Yu H. Doubly calibrated estimator for recommendation on data missing not at random. In: Proceedings of the ACM Web Conference 2024; 2024 May 13–17; Singapore. p. 3810–20. doi:10.1145/3589334.3645617.
21. Song Z, Chen J, Zhou S, Shi Q, Feng Y, Chen C, et al. CDR: conservative doubly robust learning for debiased recommendation. In: Proceedings of the 32nd ACM International Conference on Information and Knowledge Management; 2023 Oct 21–25; Birmingham, UK. p. 2321–30. doi:10.1145/3583780.3614805.

22. Allison PD. Missing data. V 23. Thousand Oaks, CA, USA: Sage; 2009. p. 72–89.
23. Newman DA. Missing data: five practical guidelines. Organ Res Meth. 2014;17(4):372–411. doi:10.1177/1094428114548590.
24. Schafer JL, Graham JW. Missing data: our view of the state of the art. Psychol Meth. 2002;7(2):147–77. doi:10.1037/1082-989x.7.2.147.
25. Rubin DB. Inference and missing data. Biometrika. 1976;63(3):581–92. doi:10.1093/biomet/63.3.581.
26. Little RJA, Rubin DB. Statistical analysis with missing data. New York, NY, USA: John Wiley & Sons; 2002. doi:10.1002/9781119013563.
27. Acuña E, Rodriguez C. The treatment of missing values and its effect on classifier accuracy. In: Classification, Clustering, and Data Mining Applications: Proceedings of the Meeting of the International Federation of Classification Societies (IFCS), Illinois Institute of Technology; 2004 Jul 15–18; Chicago, IL, USA. Berlin/Heidelberg, Germany: Illinois Institute of Technology; 2004. p. 639–47. doi:10.1007/978-3-642-17103-1_60.
28. Zhang Z. Missing data imputation: focusing on single imputation. Ann Transl Med. 2016;4(1):9. doi:10.3978/j.issn.2305-5839.2015.12.38.
29. Little RJ, Rubin DB. Single imputation methods. In: Statistical analysis with missing data. New York, NY, USA: John Wiley & Sons; 2002. p. 59–74.
30. Dempster AP, Laird NM, Rubin DB. Maximum likelihood from incomplete data via the *EM* algorithm. J R Stat Soc Ser B Stat Methodol. 1977;39(1):1–22. doi:10.1111/j.2517-6161.1977.tb01600.x.
31. Rubin DB. Multiple imputation. In: Flexible imputation of missing data. 2nd ed. Boca Raton, FL, USA: Chapman and Hall/CRC; 2012. doi:10.1201/b11826-4.
32. Allison PD. Multiple imputation for missing data: a cautionary tale. Sociol Methods Res. 2000;28(3):301–9. doi:10.1177/0049124100028003003.
33. Xia R, Liu H, Li A, Liu X, Zhang Y, Zhang C, et al. Incomplete graph learning: a comprehensive survey. Neural Netw. 2025;190:107682. doi:10.1016/j.neunet.2025.107682.
34. Singh M. Scalability and sparsity issues in recommender datasets: a survey. Knowl Inf Syst. 2020;62(1):1–43. doi:10.1007/s10115-018-1254-2.
35. Najafabadi MK, Mahrin MN. A systematic literature review on the state of research and practice of collaborative filtering technique and implicit feedback. Artif Intell Rev. 2016;45(2):167–201. doi:10.1007/s10462-015-9443-9.
36. Jalili M, Ahmadian S, Izadi M, Moradi P, Salehi M. Evaluating collaborative filtering recommender algorithms: a survey. IEEE Access. 2018;6:74003–24. doi:10.1109/access.2018.2883742.
37. Koren Y, Bell R, Volinsky C. Matrix factorization techniques for recommender systems. Computer. 2009;42(8):30–7. doi:10.1109/mc.2009.263.
38. Konstan JA, Miller BN, Maltz D, Herlocker JL, Gordon LR, Riedl J. GroupLens: applying collaborative filtering to Usenet news. Commun ACM. 1997;40(3):77–87. doi:10.1145/245108.245126.
39. Sarwar B, Karypis G, Konstan J, Riedl J. Item-based collaborative filtering recommendation algorithms. In: Proceedings of the 10th International Conference on World Wide Web; 2001 May 1–5; Hong Kong, China. p. 285–95. doi:10.1145/371920.372071.
40. Kurucz M, Benczúr AA, Csalogány K. Methods for large scale SVD with missing values. In: Proceedings of KDD Cup and Workshop. San José, CA, USA: Citeseer; 2007. Vol. 12, p. 31–8.
41. Brand M. Incremental singular value decomposition of uncertain data with missing values. In: Computer Vision—ECCV 2002. Berlin/Heidelberg, Germany: Springer; 2002. p. 707–20. doi:10.1007/3-540-47969-4_47.
42. Munson J, Cummins B, Zosso D. An introduction to collaborative filtering through the lens of the Netflix Prize. Knowl Inf Syst. 2025;67(4):3049–98. doi:10.1007/s10115-024-02315-z.
43. Panda DK, Ray S. Approaches and algorithms to mitigate cold start problems in recommender systems: a systematic literature review. J Intell Inf Syst. 2022;59(2):341–66. doi:10.1007/s10844-022-00698-5.
44. Chen J, Dong H, Wang X, Feng F, Wang M, He X. Bias and debias in recommender system: a survey and future directions. ACM Trans Inf Syst. 2023;41(3):1–39. doi:10.1145/3564284.

45. Li H, Zheng C, Wang S, Wu K, Wang E, Wu P, et al. Relaxing the accurate imputation assumption in doubly robust learning for debiased collaborative filtering. In: Proceedings of the Forty-First International Conference on Machine Learning; 2024 Jul 21–27; Vienna, Austria.
46. Adomavicius G, Kwon Y. New recommendation techniques for multicriteria rating systems. IEEE Intell Syst. 2007;22(3):48–55. doi:10.1109/mis.2007.58.
47. Monti D, Rizzo G, Morisio M. A systematic literature review of multicriteria recommender systems. Artif Intell Rev. 2021;54(1):427–68. doi:10.1007/s10462-020-09851-4.
48. Zheng Y, Wang DX. Multi-criteria decision making and recommender systems. In: Proceedings of the 28th International Conference on Intelligent User Interfaces; 2023 Mar 27–31; Sydney, Australia. p. 181–4. doi:10.1145/3581754.3584163.
49. Zheng Y. Utility-based multi-criteria recommender systems. In: Proceedings of the 34th ACM/SIGAPP Symposium on Applied Computing; 2019 Apr 8–12; Limassol, Cyprus. p. 2529–31. doi:10.1145/3297280.3297641.
50. Hong M, Jung JJ. Multi-criteria tensor model for tourism recommender systems. Expert Syst Appl. 2021;170:114537. doi:10.1016/j.eswa.2020.114537.
51. Zheng Y. Criteria chains: a novel multi-criteria recommendation approach. In: Proceedings of the 22nd International Conference on Intelligent User Interfaces; 2017 Mar 13–16; Limassol, Cyprus. p. 29–33. doi:10.1145/3025171.3025215.
52. Nilashi M, Salahshour M, Ibrahim O, Abbas M, Esfahani MD, Zakuan N. A new method for collaborative filtering recommender systems: the case of Yahoo! movies and tripadvisor datasets. J Soft Comput Decis Support Syst. 2016;3(5):44.
53. Zheng Y. Opentable data with multi-criteria ratings. arXiv:2501.03072. 2024.
54. Hong M, Jung J. Hypothetical tensor-based multi-criteria recommender system for new users with partial preferences. Comput Sci Inf Syst. 2021;18(1):285–301. doi:10.2298/csis200531056h.
55. Wang H, Lu Y, Zhai C. Latent aspect rating analysis on review text data: a rating regression approach. In: Proceedings of the 16th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining; 2010 Jul 25–28; Washington, DC, USA. p. 783–92. doi:10.1145/1835804.1835903.
56. Musto C, de Gemmis M, Semeraro G, Lops P. A multi-criteria recommender system exploiting aspect-based sentiment analysis of users' reviews. In: Proceedings of the Eleventh ACM Conference on Recommender Systems; 2017 Aug 27–31; Como, Italy. p. 321–5. doi:10.1145/3109859.3109905.
57. Cheng Z, Ding Y, Zhu L, Kankanhalli M. Aspect-aware latent factor model: rating prediction with ratings and reviews. In: Proceedings of the 2018 World Wide Web Conference on World Wide Web—WWW'18; 2018 Apr 23–27; Lyon, France. p. 639–48. doi:10.1145/3178876.3186145.
58. Zhuang Y, Kim J. A BERT-based multi-criteria recommender system for hotel promotion management. Sustain Switz. 2021;13(14):8039. doi:10.3390/su13148039.
59. Li P, Tuzhilin A. Learning latent multi-criteria ratings from user reviews for recommendations. IEEE Trans Knowl Data Eng. 2022;34(8):3854–66. doi:10.1109/tkde.2020.3030623.
60. Hong M, Jung JJ. ClustPTF: clustering-based parallel tensor factorization for the diverse multi-criteria recommendation. Electron Commer Res Appl. 2021;47:101041. doi:10.1016/j.elerap.2021.101041.
61. Ben Schafer J, Konstan JA, Riedl J. Meta-recommendation systems: user-controlled integration of diverse recommendations. In: Proceedings of the Eleventh International Conference on Information and Knowledge Management; 2002 Nov 4–9; McLean, VA, USA. p. 43–51. doi:10.1145/584792.584803.
62. Schafer JB. DynamicLens: a dynamic user-interface for a meta-recommendation system. In: Proceedings of the Wokshop Beyond Personalization 2005, in Conjunction with the International Conference on Intelligent User Interfaces IUI'05; 2005 Jan 9; San Diego, CA, USA. p. 72–6.
63. Zhang Z, Li C, Chen X, Xie X, Yu PS. Meta recommendation with robustness improvement. IEEE Trans Knowl Data Eng. 2025;37(2):781–93. doi:10.1109/tkde.2024.3509416.
64. Cantador I, Fernández-Tobías I, Berkovsky S, Cremonesi P. Cross-domain recommender systems. In: Recommender systems handbook. Boston, MA, USA: Springer; 2015. p. 919–59. doi:10.1007/978-1-4899-7637-6_27.

65. Khan MM, Ibrahim R, Ghani I. Cross domain recommender systems: a systematic literature review. ACM Comput Surv. 2018;50(3):1–34. doi:10.1145/3073565.
66. Shi Y, Larson M, Hanjalic A. Collaborative filtering beyond the user-item matrix: a survey of the state of the art and future challenges. ACM Comput Surv. 2014;47(1):1–45. doi:10.1145/2556270.
67. Schnabel T, Swaminathan A, Singh A, Chandak N, Joachims T. Recommendations as treatments: debiasing learning and evaluation. In: Proceedings of the 33rd International Conference on Machine Learning; 2016 Jun 19–24; New York, NY, USA. p. 1670–9.
68. Saito Y, Yaginuma S, Nishino Y, Sakata H, Nakata K. Unbiased recommender learning from missing-not-at-random implicit feedback. In: Proceedings of the 13th International Conference on Web Search and Data Mining; 2020 Feb 3–7; Houston, TX, USA. p. 501–9. doi:10.1145/3336191.3371783.
69. Bonner S, Vasile F. Causal embeddings for recommendation. In: Proceedings of the 12th ACM Conference on Recommender Systems; 2018 Oct 2–7; Vancouver, BC, Canada. p. 104–12. doi:10.1145/3240323.3240360.
70. Luo H, Zhuang F, Xie R, Zhu H, Wang D, An Z, et al. A survey on causal inference for recommendation. Innovation. 2024;5(2):100590. doi:10.1016/j.xinn.2024.100590.
71. Wu L, Zheng Z, Qiu Z, Wang H, Gu H, Shen T, et al. A survey on large language models for recommendation. World Wide Web. 2024;27(5):60. doi:10.1007/s11280-024-01291-2.
72. Zhao Z, Fan W, Li J, Liu Y, Mei X, Wang Y, et al. Recommender systems in the era of large language models (LLMs). IEEE Trans Knowl Data Eng. 2024;36(11):6889–907. doi:10.1109/tkde.2024.3392335.
73. Zhang L, Liu P, Deldjoo Y, Zheng Y, Gulla JA. Understanding language modeling paradigm adaptations in recommender systems: lessons learned and open challenges. arXiv:2404.03788. 2024.
74. Ifada N. Impact of imputation on cluster-based collaborative filtering approach for recommendation system. Kursor. 2019;10(1). doi:10.28961/kursor.v10i1.201.
75. Bhushan Mada SP, Tata R, Sree Reddy Thondapu ST, Saleti S. Optimizing recommendation systems: analyzing the impact of imputation techniques on individual and group recommendation systems. In: Proceedings of the 2024 IEEE International Conference on Signal Processing, Informatics, Communication and Energy Systems (SPICES); 2024 Sep 20–22; Kottayam, India. p. 1–6. doi:10.1109/spices62143.2024.10779628.
76. Ranjbar M, Moradi P, Azami M, Jalili M. An imputation-based matrix factorization method for improving accuracy of collaborative filtering systems. Eng Appl Artif Intell. 2015;46(3):58–66. doi:10.1016/j.engappai.2015.08.010.
77. Insuwan W, Suksawatchon U, Suksawatchon J. Improving missing values imputation in collaborative filtering with user-preference genre and singular value decomposition. In: Proceedings of the 2014 6th International Conference on Knowledge and Smart Technology (KST); 2014 Jan 30–31; Chonburi, Thailand. p. 87–92. doi:10.1109/kst.2014.6775399.
78. Srebro N, Jaakkola T. Weighted low-rank approximations. In: Proceedings of the 20th International Conference on Machine Learning (ICML-03); 2003 Aug 21–24; Washington, DC, USA. p. 720–7.
79. Pan R, Yang T, Cao J, Lu K, Zhang Z. Missing data imputation by K nearest neighbours based on grey relational structure and mutual information. Appl Intell. 2015;43(3):614–32. doi:10.1007/s10489-015-0666-x.
80. Ren Y, Li G, Zhang J, Zhou W. The efficient imputation method for neighborhood-based collaborative filtering. In: Proceedings of the 21st ACM International Conference on Information and Knowledge Management; 2012 Oct 29–Nov 2; Maui, HI, USA. p. 684–93. doi:10.1145/2396761.2396849.
81. Xia W, He L, Gu J, He K, Ren L. Boosting collaborative filtering based on missing data imputation using item's genre information. In: Proceedings of the 2009 2nd IEEE International Conference on Computer Science and Information Technology; 2009 Aug 8–11; Beijing, China. p. 332–6. doi:10.1109/iccsit.2009.5234936.
82. Xia W, He L, Gu J, He K. Effective collaborative filtering approaches based on missing data imputation. In: Proceedings of the 2009 Fifth International Joint Conference on INC, IMS and IDC; 2009 Aug 25–27; Seoul, Republic of Korea.
83. Hwang WS, Li S, Kim SW, Lee K. Data imputation using a trust network for recommendation. In: Proceedings of the 23rd International Conference on World Wide Web; 2014 Apr 7–11; Seoul, Republic of Korea. p. 299–300. doi:10.1145/2567948.2577363.

84. Hwang WS, Li S, Kim SW, Lee K. Data imputation using a trust network for recommendation via matrix factorization. Comput Sci Inf Syst. 2018;15(2):347–68. doi:10.2298/csis170820003h.
85. Alghamedy F, Zhang J. Imputation strategies for cold-start users in NMF-based recommendation systems. In: Proceedings of the 2019 3rd International Conference on Information System and Data Mining; 2019 Apr 6–8; Houston, TX, USA. p. 119–28. doi:10.1145/3325917.3325933.
86. Collins LM, Schafer JL, Kam CM. A comparison of inclusive and restrictive strategies in modern missing data procedures. Psychol Methods. 2001;6(4):330–51. doi:10.1037/1082-989x.6.4.330.
87. Chen J, Dong H, Qiu Y, He X, Xin X, Chen L, et al. AutoDebias: learning to debias for recommendation. In: Proceedings of the 44th International ACM SIGIR Conference on Research and Development in Information Retrieval; 2021 Jul 11–15; Virtual. p. 21–30. doi:10.1145/3404835.3462919.
88. Wang J, Li H, Zhang C, Liang D, Yu E, Ou W, et al. CounterCLR: counterfactual contrastive learning with non-random missing data in recommendation. In: Proceedings of the 2023 IEEE International Conference on Data Mining (ICDM); 2023 Dec 1–4; Shanghai, China. p. 1355–60. doi:10.1109/icdm58522.2023.00174.
89. Lee Y, Kim SW, Park S, Xie X. How to impute missing ratings? Claims, solution, and its application to collaborative filtering. In: Proceedings of the 2018 World Wide Web Conference on World Wide Web—WWW'18; 2018 Apr 23–27; Lyon, France. p. 783–92. doi:10.1145/3178876.3186159.
90. Malitesta D, Rossi E, Pomo C, Di Noia T, Malliaros FD. Training-free graph-based imputation of missing modalities in multimodal recommendation. IEEE Trans Knowl Data Eng. 2026;38(5):3250–63. doi:10.1109/tkde.2026.3667005.
91. Ding Z, Tian J, Wang Z, Zhao J, Li S. Data imputation using large language model to accelerate recommender system. In: Proceedings of the 2nd EARL Workshop on Evaluating and Applying Recommender Systems with Large Language Models at ACM RecSys; 2025 Sep 22–26; Prague, Czech Republic.
92. Gondara L, Wang K. *MIDA*: multiple imputation using denoising autoencoders. In: Advances in knowledge discovery and data mining. Cham, Switzerland: Springer International Publishing; 2018. p. 260–72. doi:10.1007/978-3-319-93040-4_21.
93. Ipsen NB, Mattei PA, Frellsen J. Not-MIWAE: deep generative modelling with missing not at random data. In: Proceedings of the 9th International Conference on Learning Representations; 2021 May 3–7; Virtual.
94. Liang D, Corneli M, Latouche P, Bouveyron C. Missing rating imputation based on product reviews via deep latent variable models. In: Proceedings of the ICML Workshop on the Art of Learning with Missing Values; 2020 Jul 17; Virtual.
95. Zheng Y. ChatGPT for teaching and learning: an experience from data science education. In: Proceedings of the 24th Annual Conference on Information Technology Education; 2023 Oct 11–14; Marietta, GA, USA. p. 66–72. doi:10.1145/3585059.3611431.
96. Xu H, Gan W, Qi Z, Wu J, Yu PS. Large language models for education: a survey. arXiv:2405.13001. 2024.
97. Zheng Y, Lu X, Yu X, Kambhampati V. Personalizing educational responses with LLMs: the influence of knowledge, interests, and preferences. In: Proceedings of the 26th ACM Annual Conference on Cybersecurity & Information Technology Education; 2025 Nov 6–8; Sacramento, CA, USA. p. 141–7. doi:10.1145/3769694.3771123.
98. Li Y, Wang S, Ding H, Chen H. Large language models in finance: a survey. In: Proceedings of the 4th ACM International Conference on AI in Finance; 2023 Nov 27–29; Brooklyn, NY, USA. p. 374–82. doi:10.1145/3604237.3626869.
99. Zheng Y, Zhang J, Shukla KN, O'Leary M, Wang DX, Xu J. Leveraging interactive visualizations and LLM-driven explanations for transparent multi-objective portfolio management. Discover Data. 2025;3(1):50. doi:10.1007/s44248-025-00065-z.
100. Wang DX, Zheng Y, Charney J. TSRMTGen: leveraging LLMs for financial narrative summaries from risk models. In: Trends and applications in knowledge discovery and data mining. Singapore: Springer Nature; 2026. p. 351–6. doi:10.1007/978-981-92-2014-4_28.
101. Al Nazi Z, Peng W. Large language models in healthcare and medical domain: a review. Informatics. 2024;11(3):57. doi:10.3390/informatics11030057.
102. Yang R, Tan TF, Lu W, Thirunavukarasu AJ, Ting DSW, Liu N. Large language models in health care: development, applications, and challenges. Health Care Sci. 2023;2(4):255–63. doi:10.1002/hcs2.61.

103. Wang Q, Li J, Wang S, Xing Q, Niu R, Kong H, et al. Towards next-generation LLM-based recommender systems: a survey and beyond. arXiv:2410.19744. 2024.
104. Boka TF, Niu Z, Neupane RB. A survey of sequential recommendation systems: techniques, evaluation, and future directions. Inf Syst. 2024;125(5):102427. doi:10.1016/j.is.2024.102427.
105. Quadrana M, Cremonesi P, Jannach D. Sequence-aware recommender systems. ACM Comput Surv. 2019;51(4):1–36. doi:10.1145/3190616.
106. Deldjoo Y, He Z, McAuley J, Korikov A, Sanner S, Ramisa A, et al. A review of modern recommender systems using generative models (gen-RecSys). In: Proceedings of the 30th ACM SIGKDD Conference on Knowledge Discovery and Data Mining; 2024 Aug 25–29; Barcelona, Spain. p. 6448–58. doi:10.1145/3637528.3671474.
107. Rajput S, Mehta N, Singh A, Keshavan RH, Vu T, Heldt L, et al. Recommender systems with generative retrieval. Adv Neural Inf Process Syst. 2023;36:10299–315. doi:10.52202/075280-0452.
108. Fan Z, Xu K, Dong Z, Peng H, Zhang J, Yu PS. Graph collaborative signals denoising and augmentation for recommendation. In: Proceedings of the 46th International ACM SIGIR Conference on Research and Development in Information Retrieval; 2023 Jul 23–27; Taipei, Taiwan. p. 2037–41. doi:10.1145/3539618.3591994.
109. Liu P, Zhang L, Gulla JA. Pre-train, prompt, and recommendation: a comprehensive survey of language modeling paradigm adaptations in recommender systems. Trans Assoc Comput Linguist. 2023;11(3):1553–71. doi:10.1162/tacl_a_00619.
110. Nnanna P, Amujo O, Ezenkwu CP, Ibeke E. Leveraging LLMs for user rating prediction from textual reviews: a hospitality data annotation case study. Information. 2025;16(12):1059. doi:10.3390/info16121059.
111. Ahn S, Shin S, Seo YD. Enriching semantic profiles into knowledge graph for recommender systems using large language models. In: Proceedings of the 32nd ACM SIGKDD Conference on Knowledge Discovery and Data Mining V. 1; 2026 Aug 9–13; Jeju Island, Republic of Korea. p. 25–36.
112. Zheng Y, Zhang J. OmniSim: a LLM-powered open-source simulator for generating personalized and adaptive conversational recommendation dialogues. In: Proceedings of the 34th ACM Conference on User Modeling, Adaptation and Personalization; 2026 Jun 8–11; Gothenburg, Sweden. p. 525–7. doi:10.1145/3774935.3812730.
113. Liang T, Jin C, Wang L, Fan W, Xia C, Chen K, et al. LLM-REDIAL: a large-scale dataset for conversational recommender systems created from user behaviors with LLMs. In: Proceedings of the Findings of the Association for Computational Linguistics ACL 2024; 2024 Aug 11–16; Bangkok, Thailand. p. 8926–39. doi:10.18653/v1/2024.findings-acl.529.
114. Feng Y, Liu S, Xue Z, Cai Q, Hu L, Jiang P, et al. A large language model enhanced conversational recommender system. arXiv:2308.06212. 2023.
115. Kim WS, Lim S, Kim GW, Choi SM. Extracting implicit user preferences in conversational recommender systems using large language models. Mathematics. 2025;13(2):221. doi:10.3390/math13020221.
116. Yuan W, Nguyen QVH, He T, Chen L, Yin H. Manipulating federated recommender systems: poisoning with synthetic users and its countermeasures. In: Proceedings of the 46th International ACM SIGIR Conference on Research and Development in Information Retrieval; 2023 Jul 23–27; Taipei, Taiwan. p. 1690–9. doi:10.1145/3539618.3591722.
117. Wang Z, Yu J, Gao M, Yuan W, Ye G, Sadiq S, et al. Poisoning attacks and defenses in recommender systems: a survey. arXiv:2406.01022. 2024.
118. Zheng Y, Mobasher B. Context-aware recommendations. In: Collaborative recommendations: algorithms, practical challenges and applications. Singapore: World Scientific Publishing; 2018. p. 173–202. doi:10.1142/9789813275355_0005.
119. Zheng Y. Context-aware collaborative filtering using context similarity: an empirical comparison. Information. 2022;13(1):42. doi:10.3390/info13010042.
120. Zheng Y. Non-dominated differential context modeling for context-aware recommendations. Appl Intell. 2022;52(5):5315–34. doi:10.1007/s10489-021-03027-5.
121. Cheng C, Yang H, Lyu MR, King I. Where you like to go next: successive point-of-interest recommendation. Int Jt Conf Artif Intell. 2013;13:2605–11.
122. Zhang Q, Yang P, Yu J, Wang H, He X, Yiu SM, et al. A survey on point-of-interest recommendation: models, architectures, and security. IEEE Trans Knowl Data Eng. 2025;37(6):3153–72. doi:10.1109/tkde.2025.3551292.